

Europe's neglect of its own future

President Bill Clinton's trip to Europe last week has raised important questions about Europe's apparent indifference to the problems of Central Europe and Yugoslavia.

PRESIDENT Bill Clinton is probably still pleased with his journey to Europe last week. For one thing, he is the first US president to have visited Minsk, the capital of Belarus; history cannot take that distinction away from him. His demonstration to the Russian people on television of how even politicians can be open in their dealings with random interlocutors will also leave a lasting impression. It is less encouraging that Mr Yegor Gaidar, the bastion of reform in President Boris Yeltsin's government, should have resigned his cabinet post the day after Clinton left Moscow. (Luckily, Gaidar remains a member of the Russian parliament.) And the signature in Moscow of an agreement between the United States on the one hand and Russia and the Ukraine on the other that the Ukraine will transfer ex-Soviet nuclear weapons to Russia over the next decade and then benefit from a rumoured \$12 billion may yet prove a dead letter; the president of the Ukraine has yet to win the consent of his unruly parliament in Kiev and, anyway, he needs a higher price.

But surely, Clintonites will ask, the "Partnership for Peace" the president proclaimed at the meeting of the North Atlantic Treaty Organization (NATO) in Brussels last week is here to stay? Certainly, Clinton put across the notion well, so well that politicians in Warsaw, Prague, Bratislava and Budapest spoke of their pleasure at what they had been promised. Of course, they would have preferred outright membership of NATO for the sake of the explicit military guarantees that that would automatically have given them, but Clinton could offer only a modicum of military liaison and the promise of stronger links at some time in the future. In other words, the nature of the promise is far from clear.

Time passes. Five years have gone since the Berlin Wall came down and since people on both sides of it were enlivened by an extraordinary sense of hope. Peace and prosperity were to be Europe's future. So why should the states of Central Europe now be alarmed about their security? Because prosperity has proved to be cruelly elusive. The West, and Western Europe in particular, could have helped decisively to right that state of affairs by more imaginative policies, notably by enabling Central Europe to trade its way out of its immediate difficulties. But that chance was missed. Now, even the hope of peace is put in hazard by the signs that the Ukraine and Russia are dangerously unstable. It is natural that Central Europe, denied assistance on the scale that might have helped five years ago, is now clamouring for security guarantees from NATO,

only to be told that it may have them at some stage in the future. That is a measure of the West's failure.

It is not too late to put things right. The obvious institution with which to start is not NATO, but the European Union (EU), now planning to enlarge itself by the accession of Austria, Finland and Sweden, but which has not yet even the sketchiest timetable for the formal membership of the Central European states. Why the delay? There are difficulties on both sides. Although the Central European states have come a long way in five years, persisting traces of the old command economies are incompatible with the single market, while the new democracies have not yet shown themselves to be constitutionally robust. In the West, the integration of Central Europe into the EU would require, under present rules, an eastward transfer of resources that would disturb the comfort (already shaken by the continuing recession) to which Western Europe believes it is entitled. But that is a poor excuse for inaction. If, after another five years, the EU discovers that Central Europe has drifted back into its old sphere of influence, it will bitterly regret the lost opportunity. Should it not have the courage to seize it now?

Ex-Yugoslavia, and Bosnia in particular, is another matter. Under pressure from France and the United States, NATO agreed last week that there may be air-strikes against aircraft and ground troops that violate negotiated agreements, which reaffirms the position reached last year — and leaves last year's objections to this course of action unchanged. Clinton was right last week to complain that Western Europe's standing has been undermined by three years of Balkan warfare, but NATO will not be the instrument of its solution. Nor can it help much if mediators from the EU and the United Nations continue their patient redrawing of the map of Bosnia to accommodate the most recent territorial gains and losses. If the EU had the wit, it would anticipate, before much more blood is shed, the political framework within which a settlement will eventually have to be reached not just in Sarajevo but in Belgrade as well. □

Defection from EMBL

Italy's plan to leave the international Heidelberg laboratory should prompt reconsideration of its function.

ITALY's decision to leave the European Molecular Biology Laboratory (EMBL) is not the end of the world for the



laboratory which is in Heidelberg in Germany, but is a serious setback. Moreover, the schemes being canvassed to win Italy back into the fold raise questions about the long-term function of the laboratory that are certain to give pause to other member states. That is why it is important that EMBL should not be confused with the European Molecular Biology Organization (EMBO), also based at Heidelberg, which runs workshops and offers fellowships throughout Europe, and which publishes the excellent *EMBO Journal*. EMBL, founded in 1974, is an offshoot of EMBO, but is constitutionally independent of the organization.

The case for EMBL was controversial from the outset. The model was that of the European Laboratory for Particle Physics (CERN) in Geneva, whose existence is predicated on the belief that it can take on projects too big for individual members states to tackle. Constitutionally, EMBL thus owes its existence to a treaty between the members states, whose annual contributions to the cost are fixed by formula. At the beginning (in the early 1970s), there were cogent objections from potential members that molecular biology differs from high-energy physics in being a relatively small-scale enterprise and not so much a field of study in its own right, as a technique destined to spread throughout biology. But there were counter-arguments: Twenty years ago, the nucleotide sequence of *E. coli* was commonly offered as a large scale goal, as was the study of biomolecular structure. Now, EMBL is Europe's custodian of the huge volume of sequence data now accumulating, its outstation at the Synchrotron Laboratory at Grenoble a symbol of its commitment to structures.

But the old objection persists. If anything, molecular biology has turned out to be more pervasively a part of biology in general than even the critics imagined twenty years ago. That is not to say that EMBL's core laboratory at Heidelberg serves no purpose. On the contrary, it maintains several energetic research groups which, because of their international composition, have an incidentally beneficial effect on the spread of molecular biology through Europe. The benefits may be expensive, but at least they are tangible. If Italy pulls out, the others should be prepared to soldier on at least until Dr George Kafatos, the new director, has a chance to settle in.

The hope of keeping Italy sweet by offering to set up outstations in Italy (see page 205) is, on the other hand, misguided. No doubt part of the Italian suspicion that it derives less benefit than others from the present set-up centres around last year's decision to transfer the sequence databases to a new centre at Cambridge, England (with the help of Wellcome money), but Italy's commitment to the International Centre of Genetic Engineering and Biotechnology at Trieste may also have influenced the decision. But the plain truth is that the proliferation of outstations would further weaken EMBL's claim on public attention, turning it from a research organization into a way of recycling national funds into internationally supported national laboratories. What EMBL needs is a research programme, not a funding mechanism. □

The price is wrong!

Eurotunnel's high prices for crossing the English Channel by motor car are disappointingly high.

THE legend of the Channel Tunnel (connecting Britain and France) has its roots in the eighteenth century. When tunnelers were skilled (but not particularly skilled) at cutting tunnels only through coal-seams, and when the standard means of traction was the horse, the project could have been laughed out of court on technical grounds, but was not; the British fear that Napoleon would march his armies beneath the Channel (otherwise *La Manche*) was a more potent reason for abandoning the project. After the Congress of Vienna brought peace to Europe and with the arrival of steam locomotion, the project was revived, a few hundred metres of tunnel were dug beneath the sea at Dover and then again abandoned when the organizers ran out of money. But now a Channel Tunnel exists and will open for service in a month or so. The first passengers will travel with their motor cars, but a passenger train service will begin in June or thereabouts. So where are the signs of general rejoicing such a development should command?

Sadly, there is very little rejoicing, and what there is has been dampened by the announcement last week by Eurotunnel, the Anglo-French company franchised to operate the tunnel, of its tariffs for carrying motor vehicles beneath the Channel. Like the ship ferries that now ply the Channel, Eurotunnel plans different charges at different times of the year. Broadly speaking, Eurotunnel will charge drivers a little less than the ferry companies to cross the Channel at peak summer weekends (£310 against £320), but considerably more at other times (£220 against roughly £130). For Eurotunnel, that seems a recipe for an awkward seasonal fluctuation of demand; novelty and the slightly lower price will ensure that its summer weekends are busy, but the ferries will take the lion's share of the traffic at other times. That seems unwise when the business hangs on loading motor cars onto untried transporters — but luckily, that is Eurotunnel's business, not that of either British or French taxpayers.

The more serious disappointment is that the fares are so high. The reasons are plain enough; Eurotunnel was originally meant to cost £4.8 billion, but the final cost will be more than twice as much. At these rates, the Channel ports from which the ferries operate can sleep easily again, confident that Eurotunnel will capture a smaller part of the cross-Channel vehicle business than they had feared. And that is a misfortune for others on both sides of the Channel because Europe is much in need of a way of making Britain more a part of the rest of Europe. The only hope now is that the cost of travelling through Eurotunnel by train, still to be declared, can be kept substantially below the air fares charged by the rapacious airlines operating between Britain and the mainland. That way, some social benefit may be won from the investment of all the technology that has gone into Eurotunnel. □

Italy to quit European biology lab in protest over staffing policy

Bonn. Italy has shocked the world of European science by announcing its intention to resign from the European Molecular Biology Laboratory (EMBL), based at Heidelberg in Germany, claiming that it does not get a fair return on its annual subscription to the laboratory.

The decision was announced on 28 December, and will, unless reversed, come into effect at the end of this year. It has surprised many EMBL council members, who had previously been optimistic that Italy's longstanding threat to withdraw had been averted by a new plan to set up four regional research groups in Italy (see *Nature* 366, 604; 1993).

The decision also throws EMBL's plans for the next few years into chaos. Italy contributes around 16 per cent of EMBL's annual budget of about DM10 million. Its withdrawal will therefore be a major blow at a time when the laboratory is expanding two of its three outstations, in Grenoble (France) and Cambridge (United Kingdom).

But laboratory officials still hope that Italy can be persuaded to change its mind. Italy first threatened to pull out of EMBL

early last year, because it believed it was underrepresented among EMBL staff scientists. Its resignation letter indicates that it would be prepared to rejoin if the conditions were appropriate.

EMBL was founded in 1974 as a European focal point for molecular biology, and is currently paid for by subscriptions from 15 member states. It has always opposed a quota system for appointing staff scientists, arguing that appointments should be made on scientific merit alone.

The laboratory's new director, Fotis Kafatos, who was appointed last April, has suggested an alternative solution to complaints that this policy discriminates against member states that are less strong in molecular biology, namely that regional groups should be set up throughout southern European member states.

Under Kafatos's proposal, four of the five groups planned for a pilot scheme to start this year would be located in Italy, at a cost of about DM450,000 each. In a second phase of the pilot scheme, a further six would be established in Italy over the next four years, to be paid for out of funds raised

	Staff Scientists		Staff Technicians		Fellows		Visiting Scientists		% of budget
	'89	'93	'89	'93	'89	'93	'89	'93	
Italy	9	4	4	3	24	21	5	7	16.2
Germany	30	25	50	51	32	55	5	14	22.0
UK	20	26	32	44	29	17	7	17	14.4
France	9	9	16	21	18	26	3	8	17.0

Losing out: Italy claims its scientists are underrepresented on EMBL's staff.

from outside bodies such as the European Commission.

The scheme was enthusiastically endorsed at EMBL's annual council meeting on 9 December. At the time, Italian officials seemed to regard it as an acceptable resolution of their grievances. There was therefore both surprise and disappointment when, less than three weeks later, Italy announced its withdrawal, saying it wanted to stimulate debate about the "national and international" reasons that have led to its low representation among the staff of EMBL facilities.

The number of Italian staff at EMBL has fallen considerably in the last few years, coinciding with an expansion of molecular biology in Italy, and the return of key scientists to take up new positions.

But the number of Italian staff scientists and technicians has always been low compared with those from Germany and the UK. Various reasons have been put forward to account for this, including the underdevelopment of molecular biology in Italy.

Arturo Falaschi, head of the International Centre of Genetic Engineering and Biotechnology in Trieste, who has been acting as an informal adviser to the Italian government, says that one factor in its decision was disappointment over what Italy sees as a lack of commitment by the council to the second phase of the pilot regional group programme.

Falaschi describes his government's action as "constructive", because it will ensure immediate discussion between EMBL members of its recruitment policy, a discussion in which Italy is keen to take part. "But we want to be in a better position to negotiate fairer returns," he says.

He adds that Italy is not demanding a full-blooded quota system, but merely some way of ensuring that EMBL employs more Italian scientists. He also suggests that greater consideration should be given to the creation of an outstation in Italy.

The decision has upset many of Italy's molecular biologists. Riccardo Cortesi, for example, who worked at EMBL for 11 years before moving to Rome in 1990 to run a major molecular biology laboratory, ►

Hubble shows off its new spectacles

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

NASA

Astronomers and space scientists have been waiting with fingers crossed to see if the first images from the repaired Hubble Space Telescope were as good as predicted. Their wait has been rewarded. The picture on the right shows the nuclear region of NGC1068, an active Type 2 Seyfert galaxy, with unprecedented clarity.

Previous observations (left) with the European Space Agency's Faint Object Camera (FOC) showed a number of hot gaseous clouds surrounding the intense nuclear source. New observations made by a team led by Duccio Macchetto of the European Space Agency and the Space Telescope Science Institute in Baltimore, Maryland, using the FOC with the corrective optics unit (COSTAR) installed by the astronauts, show a much more extensive area of emission.

The new picture also reveals filamentary detail in gas cloud near the nucleus in unprecedented detail. As the collective sigh of relief turns to jubilation, NASA is hoping for some respite from an increasingly critical public (see page 207). □

describes the withdrawal as an act of bad faith.

A petition signed by more than a hundred scientists has been sent to the minister of research, Umberto Colombo, and many individual protests have also been sent.

Kafatos says he intends to work hard over the next few months with Kari Kivirikko, the new president of the EMBL council, to try to resolve the dispute. "I deeply regret Italy's decision because I believe the association between Italy and EMBL is mutually beneficial, and also beneficial for science in Europe in general", says Kafatos, although he adds that the issues raised by Italy are "long-standing and need attention".

Plans for the regional groups are temporarily in abeyance. Kafatos says that the programme will go ahead, but exactly how, and where the groups will be located, are once again open for discussion.

Meanwhile senior scientists at EMBL's main laboratory in Heidelberg have been asked to look for potential savings in their research projects, in case Italy does not reverse its decision to resign.

The EMBL council meets next month to continue discussion of a proposal to raise contributions by two per cent in real terms to cover the costs of expanding its two outstations. No decision on this proposal was reached at last month's meeting, and the new meeting will also provide an opportunity to discuss Italy's resignation.

The outcome of this debate will be watched closely. The issues affect not only EMBL, but any joint scientific project in Europe, and Italy's action could set a precedent for other countries that feel that they do not get an adequate return for their investment in major international facilities.

Alison Abbott

Mitterrand asked to pardon doctors in HIV blood scandal

Paris. Ninety-seven researchers, physicians and public health workers this week sent a letter to President François Mitterrand asking him to pardon the four physicians convicted in July last year for their part in the contamination of French blood supplies with human immunodeficiency virus (HIV) in the mid-1980s (see *Nature* 364, 269; 1994).

The court handed down four-year prison terms to two former officials of the French National Centre for Blood Transfusion (CNTS), Michel Garretta and Jean-Pierre Allain, with two years of the sentence suspended. Robert Netter, former head of the National Health Laboratory, and Jacques Roux, former director-general of health at the National Ministry of Health, were given one- and three-year suspended sentences respectively.

In their letter to Mitterrand, the signatories — most of whom are from France, although some are from elsewhere — describe the verdicts as "unfair". They say that the trial was prejudiced by intense media attention, and by the passions aroused by the affair among the general public.

Referring to the alleged scientific uncertainty worldwide about HIV in 1985, the letter argues that the delay between scientific discoveries and their "validation, recognition and integration into general consciousness" is much longer than the public appreciates. "It is deeply unjust to punish a few selected individuals for the lacunae and dysfunctions of our health system as a whole," the letter adds.

The group says that its main concern is the possibility that those involved in making decisions "at the frontiers of scientific knowledge" could face similar verdicts in the future. The letter claims that fear of such "legal reprisals" could prevent physicians and researchers from "assuming their duty and from taking the responsibilities that are theirs", and thus hinder medical progress.

While confirming their "solidarity and compassion for the haemophiliacs and their families stricken by AIDS", the group says that it has a duty to guard against "any unjustified sentiment of mistrust against physicians and medical science, which would be counterproductive to medical health". It calls on Mitterrand to use his powers to prevent such a backlash, and asks him to grant a presidential pardon to the four convicted physicians.

The main signatories of the letter are Françoise Barré-Sinoussi of the Pasteur Institute, Paris, one of the co-discoverers of HIV, and Jean Claude Gluckmann of the Hôpital de la Pitié in Paris. Other signatories include William Blattner (National Cancer Institute, Bethesda, USA), Jean-Baptiste Brunet (European AIDS centre), Robin Carrell (Department of Haematology, Cambridge, UK), Jean Dausset (CEPH), Marc Giraud (Pasteur Institute, Paris), Willy Rozenbaum (Rothschild Hospital, Paris), Jorg Schupbach (Swiss National Centre for Retroviruses, Zurich) and Daniel Tarantola (International AIDS Program, Cambridge, USA).

Declan Butler

Asian republics agree on joint rescue plan to save the Aral Sea

London. Five central Asian republics, as well as Russia, formally agreed last week to set up a joint fund to save the inland Aral Sea, heavily depleted by the use of water for irrigating rice and cotton in surrounding regions.

Meeting in the Uzbek town of Nukus, the presidents of Uzbekistan, Kazakhstan, Kyrgyzstan, Turkmenistan, and Tajikistan, together with Russian Deputy Premier Yuriy Yarov, adopted an agreement to set up the fund, to which four of the five republics have agreed to provide one per cent of their annual state budgets.

The participants also agreed to take urgent steps to address the ecological problems of the area through the activities of an interstate council. The agreement on a conservation plan follows the decision by the World Bank in Washington and the United Nations Environment Programme to approve a grant to the five republics to carry out work on the prob-

lems of the area.

First attempts to prevent the complete disappearance of the Aral Sea were made as a result of concerns expressed through the Academy of Sciences in the 1980s. However these efforts became victims of recent political upheavals in Moscow, leading to fears that the rescue mission might have to be aborted.

The shrinking of the sea, which was once the world's fourth-largest freshwater lake, has left many ships stranded in what is now little more than a sandy desert whose atmosphere is polluted with fertilizer toxins. This has itself caused widespread health problems in the local population.

Addressing a recent meeting organized by the World Health Organization in Alma-Ata, President Nazarbayev of



Sea of troubles: without help, the Aral Sea is in danger of running completely dry.

Kazakhstan said that he had been given promises by a number of heads of state and international agencies that they would support the new fund both technologically and financially. "Its aim is first of all to help people," Nazarbayev said. □

Moon mission poses dilemma for scientists

Washington. Next Tuesday (25 January) will see the launch from the Vandenberg Air Force Base in California, of the first US mission to the Moon for 20 years. After spending several months photographing the lunar surface, the spacecraft, known as Clementine, will fly back past Earth for a high-speed encounter with the asteroid Geographos in August.

But planetary scientists are in two minds about the implications of their participation in the \$75-million mission, which is being funded jointly by the Department of Defense (DoD) and the National Aeronautics and Space Administration (NASA).

For the gathering of scientific data is a bonus on a trip originally conceived by President Ronald Reagan's Strategic Defense Initiative Office — the body set up to run the 'Star Wars' programme — now scaled down and renamed the Ballistic Missile Defense Organization.

Star Wars required lightweight, 'intelligent' spacecraft capable of hunting down fast-moving objects. As testing such a craft in Earth orbit would breach the Anti-Ballistic Missile (ABM) Treaty, Clementine was born to carry out the tests in outer space.

With the thawing of the Cold War, a group of planetary scientists, led by Eugene Shoemaker of the US Geological Survey, thought they could hitch a ride on board a military mission which had become an open secret. As a result, NASA set up a team early in 1993 under Shoemaker to plan the scientific exploitation of data from the craft, which has been built by the Naval Research Laboratory in Washington.

Unlike previous DoD or NASA spacecraft, Clementine uses advanced electronics and optics technology already used in devices such as camcorders and portable computers. The sensors designed for Clementine by the Lawrence Livermore laboratory in California use little power, and weigh less than a kilogram.

Not all planetary scientists are happy, however, about riding on Star Wars' coat tails. They are concerned less about the morality of the Star Wars programme itself — an old controversy that has been largely defused by historical events — than the impact of Clementine on the prospects for 'real' scientific missions both to asteroids and to the Moon.

"Many scientists are at least uneasy about working with military programmes in space," says Clark Chapman of the Planetary Science Institute, a division of the research company SAIC based in Tuscon, Arizona. But Chapman says that he shares more specific concerns about Clementine with a sizeable faction of planetary scientists.

"The approach that has been used definitely makes science a second-class citizen," Chapman says. "The threat is that it

will be decided that this is the right way to do business: quick, cheap [space vehicles] using off-the-shelf instruments and perhaps not returning much science."

Supporters of Clementine point out that NASA is already funding at least one 'real' asteroid mission, the Near Earth Asteroid Rendezvous (NEAR) which is under construction at Johns Hopkins University in Maryland, and will be launched to the asteroid Eros in 1996.

These scientists say that the critics of Clementine are allowing their political view of Star Wars to colour their judgement. "It's a good idea, and one which will hardly cost NASA anything," says Jeffrey Bell, an asteroid astronomer at the University of Hawaii. "Why turn down a free lunch?"

But the critics, who had previously been worried that Clementine's existence might jeopardize NEAR, are now fretting about

the future. These two missions will provide information about the surface appearance of asteroids: data about craters, for example, will help astronomers to develop models showing how many rocks may be flying around in the Solar System.

Geologists, however, want to study rock samples returned from asteroids. They worry that Clementine may delay the day they get them, because a number of smaller missions could delay the launch of a larger mission.

Such worries are probably unfounded. Clementine is primarily an engineering, not a science, mission. Experience from it will help make future missions cheaper, smaller and quicker. And the ceilings being proposed on future NASA mission costs should ensure that more probes are sent to near-Earth asteroids and the Moon, as the agency will not be able to afford to go much further afield.

Colin Macilwain

Hubble success boosts NASA's star

Washington. A delighted Hubble Space Telescope science team released the first images from the repaired telescope last week, declaring, in the words of Ed Weiler, a programme scientist, that "the Hubble is fixed beyond our wildest expectations".

Ever since taking a first look at images from the corrected telescope in late December, astronomers say they have been amazed at how well the repairs appear to have worked. Earlier in the month, astronauts working from the shuttle had replaced the telescope's Wide Field/Planetary Camera (WFPC), installed a system of mirrors called COSTAR that restores three other instruments to near-full capability, replaced solar panels and gyroscopes, and upgraded computers.

Project managers had thought they would need until late January to calibrate and adjust the newly installed instruments from the ground. But the work went faster than expected. "We've had to devote no time whatsoever to troubleshooting," says David Leckrone of the National Aeronautics and Space Administration's Goddard Space Flight Center.

Leckrone says that the handling of the instruments by the astronauts in space had been so "benign, even gentle", that the instruments were barely knocked out of their pre-launch alignment.

Two onboard spectrographs are still being adjusted. But the Hubble team is confident that they, too, have been restored to full working order. "We clearly nailed the prescription," says Weiler. Some scientific investigations with the telescope are beginning this week, and it should be back to a regular schedule of

observations by mid-February.

Team scientists say that the new WFPC images are twice as detailed as any images they could have obtained before the repair, even with computer processing to remove the effects of the flawed mirror.

Some 1,200–1,500 astronomers are waiting to use the telescope now that it has been fixed. Early priorities include spectrographic searches for evidence of black holes and searches for evidence of planets around other stars.

On the basis of the performance of the repaired telescope, scientists also are confident that they will be able to determine the rate of expansion of the Universe to an accuracy of 10 per cent.

The scientists and engineers who designed and built Hubble feel that their efforts have finally been vindicated by the new results after three and a half frustrating years. Indeed, Senator Barbara Mikulski (Democrat, Maryland), who oversees the congressional committee that funds NASA, used the occasion to declare that the space agency has shown its "right stuff" by fixing the telescope on time and on budget.

Mikulski also said that the repair will raise confidence in Congress that an orbiting space station can be successfully built and operated. Astronomers would probably prefer that any political goodwill from Hubble's triumph translates into support for two other planned space observatories — the Advanced X-Ray Astrophysics Facility (AXAF) and the Space Infrared Telescope Facility (SIRTF) — both of which face tough funding battles.

Tony Reichhardt

Nuclear centre appeals to Moscow for help

Moscow. Russia's financial difficulties are adding to growing concern over the future of the Joint Institute for Nuclear Research (JINR) at Dubna near Moscow, until now one of the few islands of stability in the stormy sea of Russian science.

The institute was set up by Josef Stalin in the late 1940s to meet the scientific needs of the COMECON countries, acting as a rival to the European Laboratory for Particle Physics (CERN) in Geneva. But in recent years many of the former communist states have found it increasingly difficult to support the research centre.

The institute has already made adjustments to meet its new circumstances. Staff numbers, for example, have been reduced from 7,000 to about 3,500. And the JINR has been turning increasingly for support to Western nations — particularly Germany — which are now making substantial contributions to its research through

collaborative agreements.

But the JINR is facing major difficulties as the result of the failure of the Russian Federation, which pays almost one-third of the institute's costs, to pay the money it owes, in particular to cover heavy electricity bills estimated at 600 million rubles (US\$500,000) a month.

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

Dubna's 10 GeV Synchrophasotron, opened in 1957.

Despite its financial difficulties, JINR continues to develop its scientific programmes, although it has shifted its attention away from high-energy physics to fields

such as the investigation of nuclear structure, operating at lower energy levels.

Vladimir Kadyshevsky, who was appointed director of the institute in June 1992, claims that the value of the institute's work is confirmed by its continued success in attracting physicists from Western countries. It has also become an important supplier to the West of radioisotopes prepared with its accelerator.

Kadyshevsky says that the agreement between the JINR and the German Ministry of Research and Technology serves as a good model of collaboration. The distribution of the German contribution to JINR's budget is laid down in an intergovernmental agreement.

Similar collaborative agreements have been negotiated with Italy and France, and are under discussion with the United States. Such agreements have helped to preserve the stability of JINR.

Kadyshevsky also points out that, unlike many other governmental or parliamentary institutions, the JINR has actively maintained its contacts to former Soviet states. "Even the Baltic states are on the verge of coming back to Dubna," he says. "Lithuania, for example, is discussing with us the possibilities of their scientists working here."

But the main difficulty now facing the JINR is the failure of Russia to make its promised payments into the institute's budget. This has become a priority concern for the organization's scientific council, made up of prominent physicists from both East and West.

The council members have written to the Russian Prime Minister, V. Chernomyrdin, and to other leading ministers — including Boris Saltykov, the minister of science and technical policy — expressing their "deep concern" about the financial situation and thus the future of the institute.

Their letter says that, despite the difficult economic circumstances, scientists at the institute have continued to produce excellent scientific results. But it also says that the failure of the Russian Federation to fulfil its obligations to JINR "threatens not only those collaborations but also the future of the institute".

In response to such appeals, First Deputy Premier Yegor Gaydar is reported to have told Boris Fedorov, the minister of finance, to find extra resources to settle the centre's total debt. According to Saltykov, an agreement in principle has also been reached on a government decision to provide scientific centres with preferential rates for electricity.

These proposals have now been submitted by Saltykov to President Boris Yeltsin for his approval. But with the present political turmoil in Russia, it is likely to be some time before the issue is resolved. **Carl Levitin**

Dubna benefits from 'brain drain'

Moscow. Although suffering from a 'brain drain' of scientists to the West, Dubna, the home of the Joint Institute for Nuclear Research (JINR), is itself benefiting from a similar process by attracting scientists from independent states where living and working conditions are even worse.

According to Alexander Rats, the town's deputy mayor, negotiations are taking place with scientists from the various regions in the former Soviet Union that have separated from Russia. Besides the JINR, Dubna is also the location of various enterprises specializing in aerospace and electronic technologies, and is keen to establish itself as a centre for high-technology research.

Dubna, like the rest of Russia, is no research paradise. Salaries are insufficient to meet minimal living needs, and are not even regularly paid. New instruments and equipment are difficult to obtain, and research institutes have heavy costs for electricity, as well as other costs unique to the Soviet scientific towns. But the situation is more stable than elsewhere, and thus has attractions for research scientists and their families.

At the end of November, for example, several leading specialists from the Sukhumi Physico-Technical Institute (SFTI) moved to the newly created Institute of Physico-Technical Problems based on the JINR site. The institute was specifically created in order to attract scientists such as those from Sukhumi, formerly one of the most beautiful Soviet resorts on the Black Sea but now heavily damaged by civil war

in independent Georgia.

Similarly, more than 20 scientists have moved to Dubna from the Institute of Radioisotopic Instrument Engineering in Riga, which faced the threat of closure since its main research topic — the development of radiation detectors — has become of little interest to the now nuclear-free Latvia.

And in a separate development, the Ukrainian aerospace design bureau 'Typhoon', having previously been forbidden to have any contact with Russian aerospace enterprises — its main source of contracts — has now moved most of its scientific staff to Dubna.

The Russian government still does not differentiate between migrants by profession, and is trying to settle everyone in the Nechernozomy region, the country's main agricultural area. Rats argues that this policy is short-sighted. "A doctor of science has no business in the Nechernozomy region," he says. "A scientist has to be provided with conditions which he can work in. The return is quick, and the expenses are not too great. The state has a duty to solve this problem."

At present, however, Dubna is using its own resources to provide for immigrants from other countries of the former Soviet Union. Scientists are offered temporary accommodation, a plot of land and a loan for building a house. But the city authorities still face the problem of obtaining residence permits for 'foreign' scientists, a process that involves much bureaucracy, and for which the town administration has to pay a substantial fee. **Vladimir Pokrovsky**

Ethics bill prompts second thoughts among scientists

Paris. Conservative politicians in the French Senate are pressing for tight restrictions on research using embryos and related aspects of human infertility. But their actions are leading to second thoughts among many who had previously supported demands for a strict legislation to ensure that the research is properly conducted and applied.

In November 1992, a broad political consensus ensured the adoption of France's three bioethics bills after their first reading in the Socialist-led national assembly. The bills were intended as the world's first comprehensive legally-binding code of bioethics, in part reflecting the French tradition of governing professional activity through legislation rather than codes of conduct.

But signs of a rethink by the left came earlier this month when François Mitterrand, the Socialist president who has long championed efforts to legislate on bioethics, said that he is now "unsure that morals could be changed by a law".

In particular, moves by the French Senate to ban research on human embryos and genetic testing before reimplantation have reopened a debate over whether medical progress should be regulated by law or by

Researchers and physicians generally agree that the bills are both needed and "generally balanced". They also welcome the government's commitment to review legislation in five-year's time.

But many are nonetheless concerned that the latest restrictions, proposed last week by the senate's Commission on Social Affairs, "go too far". Jacqueline Mandelbaum, for example, a researcher at the *in vitro* fertilization (IVF) centre at the Necker Hospital and the INSERM laboratory on Comparative Human Genetics in Paris, opposes the ban on human embryo research.

She says that such research should improve the viability of embryos used in IVF, as well as the proportion of successful implants. Research on such questions as the mechanisms by which human embryos implant themselves is also needed to improve the understanding of *in vivo* human reproduction, she argues.

Axel Kahn, director of the INSERM Laboratory of Research on Genetics and Molecular Pathology at the Cochin Institute of Molecular Genetics in Paris, and Jean Dausset, a Nobel prizewinner for medicine, are less convinced of the need for research on human embryos. But both are concerned by the move towards tighter restrictions. Kahn describes the proposed ban on embryo research as "philosophically incoherent", and argues that respecting the 'human dignity' of an embryo need not entail banning research on it.

Kahn asks why, if experiments can be done on humans of all ages — including infants — they should not also be allowed on embryos, given the same sort of safeguards. The proposed restrictions on research, he says, are "not very defensible", because they are being made on "ideological and religious grounds".

Dausset, a leading figure in recent debates about the ethics of genetics research, says that much research can be done using animal embryos rather than human ones. But he adds that decisions should still be left to researchers. "If we legislate on everything we will go too far," says Dausset. "We must give a margin of liberty to the conscience of researchers and physicians."

All three researchers say that the proposed ban on preimplantation diagnosis and selection of embryos should allow exceptions. They suggest limiting such techniques to a defined list of serious diseases, and carrying them out at specially authorized centres. Indeed, the state already permits therapeutic abortion of fetuses carrying genetic abnormalities.

One researcher who defends the ban, however, is Jacques Testart, a researcher at

How France plans to legislate for bioethics

Paris. Three bills concerned with bioethics are being discussed in the French Senate (see left) following their approval by the National Assembly and subsequent amendment by the government. The first, from the Ministry of Justice, aims 'to protect human dignity and the human race'. The bill bans eugenics, but permits gene therapy (providing it is not germline gene therapy). It also bans genetic testing (except for research, medical or judicial purposes), surrogate motherhood and patenting of parts of the human body (including the genome).

The second bill is from the Ministry of Health. This restricts *in vitro* fertilization (IVF) to living, sterile couples (a male and a female) "of reproductive age", and bans experimental interventions on embryos as well as production of embryos for any use other than IVF (the government has explicitly ruled out IVF for widows and postmenopausal women). The bill also rules that organs cannot be transplanted without explicit consent from donors — or next of kin, if the donor is dead — and that donations must be free and anonymous.

The third bill comes from the Ministry of Research and Higher Education. It amends the data protection act to allow physicians to provide epidemiology researchers with confidential information on patients, providing that patients agree and that their identity is concealed. **D. B.**



the conscience of individual physicians and researchers.

Many French scientists and physicians are concerned that the conservative-dominated senate will yield to pressure from religious lobbies for further restrictions when it debates the bills in the coming weeks.

Furthermore, the conservative majority that dominates the national assembly since last year's elections will probably endorse the ban on embryo research, and other restrictions, when it gives the bills a final reading in spring. But the government may well oppose the senate's expected proposal to forbid destruction of spare embryos — of which there are around 20,000 in France.

INSERM's laboratory of Gamete Maturation and Fertilization. Testart argues that the ease of IVF compared with therapeutic abortion means that it could be more readily used to select for sex or other characteristics. But he also agrees that selection of embryos could be permitted in a few well-defined cases.

Testart's main concern is that physicians and researchers could interpret the bills' provisions liberally. He says that the language in the bills — for example, words such as 'sterility', 'serious illness' and 'medically assisted procreation' — is ambiguous in legal terms. Indeed benign neglect of the provisions of the law has co-evolved with the French culture of governing by legislation.

While senators want to protect spare early-stage embryos from destruction, they are holding back from ruling on the status of the embryo. To decide that an eight-celled blastomere, for example, is a person would provide a logical basis for the senate commission's proposed restrictions. But it would create political and social havoc by running into conflict with France's abortion laws.

Declan Butler

Bribery case underlines Japan's bureaucracy

Tokyo. The bureaucratic stranglehold that Japan's Ministry of Education, Science and Culture holds over the country's universities has been highlighted by a bizarre case of corruption in the Japanese government. Last week, an official from the ministry was arrested on suspicion of having accepted a bribe of ¥500,000 (US\$4,500) from senior officials at a small private university.

As a case of bribery, the incident is trivial compared with the corruption in other parts of the government that is reported almost daily. This can involve payments of tens or hundreds of millions of yen to win favours from government officials.

But the case stands out because of the peculiar purpose of the alleged bribe. Toshio Yamaura, 37, deputy chief of the ministry's Higher Education Bureau, has been accused of accepting the bribe to speed up the processing of a change of name of a faculty and its departments in Sugiyama Jogakuen University, a small university in Nagoya that started out as a sewing school in 1905.

Universities throughout Japan face a crisis because the population of 18-year-olds, having peaked at just over 2 million in 1992, is now rapidly declining, and is expected to fall to 1.5 million by the end of this decade. As a result, the hundreds of universities in Japan, especially the less well known private ones, are scrambling to make themselves more attractive to the dwindling student population to avoid being forced out of business by a lack of students.

The university in Nagoya wanted to change the somewhat old-fashioned name of its 'home economics faculty', which covers studies of food, nutrition, clothing and textiles, to the more fashionable 'school of life studies'. And within the faculty it also wanted to split the department of 'clothing and textiles' into two new departments, to be known as 'social science' and 'the human environment'.

In Japan, however, such changes need the official approval of the ministry of education in Tokyo. Such approval typically takes at least one year in the case of a

department name, and two years for changes in a faculty. The president of one of Japan's leading national universities says it took him five years to re-name and re-organize several science departments in his university, and most of this time was spent clearing the ministry's procedures.

These regulations are just one element of a complex bureaucracy that extends from the ministry throughout Japan's universities, and is meticulously maintained by administrators posted to the universities from the ministry.

Faced with the imminent decline in the country's student population, officials at Sugiyama Jogakuen University, including the head of the board of trustees, Masahiro Sugiyama, seem to have been unable to accept a wait of several years. The university won approval for the name changes in

December 1990 only about six months after it had applied for them.

A few months later, according to police charges, Sugiyama and three other university officials paid the bribe to Yamaura, and subsequently took him on a sightseeing tour of the local countryside as a way of thanking him for his unusually speedy processing of their application.

Ryoko Akamatsu, Japan's education minister, who was not in office at the time of the alleged bribery, apologized at a news conference last week and said she was "astounded" by the incident. She promised that the ministry will do what it can to ensure that such an incident does not happen again. But she did not suggest that universities should be allowed to change the names of departments and faculties themselves.

David Swinbanks

Milk hormone clears final hurdle

Washington. The publication of a favourable report by the Clinton administration on the economic and social impacts of recombinant bovine somatotropin (rBST) has helped to clear the way for the Monsanto Company to begin marketing the drug to dairy farmers when a temporary ban on its sale is lifted two weeks from now.

With almost all congressional avenues for derailing the product now exhausted, it will now be up to farmers and consumers to decide whether this new aid for the dairy industry — which will appear under the trade name Posilac more than nine years after the company first applied for regulatory approval — sinks or swims.

The drug was approved for commercial use in the United States by the Food and Drug Administration (FDA) last November (see *Nature* 366, 192; 1993). But Congress imposed a 90-day ban on sales of rBST following its approval, during which time the administration agreed to undertake a study of its implications for consumers,

dairy farmers and the economy.

The report, which was released in Washington last week, paints a rosy picture of the technology, and says that consumers and dairy farmers could stand to benefit from lower milk prices and an increase in milk yields per cow of 10–20 per cent.

But Senator Russ Feingold (Democrat, Wisconsin), one of five congressmen who had requested the study, and a vocal opponent of rBST on economic grounds, says he was "appalled at the bias of the report", which, in his view, "read like it was written in the Monsanto boardroom".

Feingold says that the report plays down the effect that rBST will have on the dairy industry and on the cost to the federal government for the milk price support programme, and focuses too heavily on the impact of the ban on Monsanto and the biotechnology industry as a whole.

For Monsanto, the only company of four so far to have obtained FDA approval to market rBST in the United States, the temporary ban has meant 90 days of lost sales and the opportunity to build market share before competing products are approved. Rather than flooding the market with free samples, Thomas McDermott, a spokesman for the company, says that Monsanto has spent the time educating farmers as to the benefits of rBST and distributing information to potential customers.

With the battle shifting away from Congress, Jeremy Rifkin of the Foundation on Economic Trends says that protest will now take the form of "an old-fashioned grassroots campaign". Starting on 3 February, he says, the battle will be waged in neighbourhoods, restaurants and at the grocery stores.

Diane Gershon

Germany eases controls on gene experiments

Munich. Germany has introduced a new law covering research in genetic engineering, overcoming a threat from the left wing of the opposition Social Democrats to block its passage through the German Parliament.

The law eases procedures for the authorization of genetic experiments. In particular it reduces substantially the amount of paperwork required for no-risk and low-risk genetic experiments, long considered a powerful disincentive for the development of biotechnology and

genetic engineering in Germany.

The Social Democrats (SPD), concerned that the law would reduce public accountability, has demanded more control of genetic experiments, particularly for field experiments with genetically modified organisms which may now proceed without a public hearing (see *Nature* 365, 684; 1993). But 14 proposed amendments were overruled by their more conservative colleagues in the SPD-led *Länder*, who sit in the upper house, and who voted with the government.

Alison Abbott

South African researchers want greater role in planning

Cape Town. An initiative set up last May to look at how science should be run in South Africa in the period leading up to majority rule has come under fire from university scientists who claim that they are insufficiently represented in the discussions currently taking place.

The purpose of the initiative has been to discuss ways in which issues relating to the country's science system and future policy needs can be handled during the transition period and in a democratic South Africa.

But at a recent meeting of the Royal Society of South Africa, Dave Woods, deputy vice-chancellor for research at the University of Cape Town, expressed concern that the initiative was just a case of "the old bureaucrats and the new bureaucrats getting together". He pointed out that the members of the working group set up to develop the initiative contain only one practising scientist, Friedel Sellschop, his counterpart at the University of the Witwatersrand.

Woods' comments echo sentiments expressed previously by a meeting of black university scientists, headed by Chris Johnson, dean of science at the University of the Western Cape. The meeting issued a statement calling for a moratorium on "all unilateral restructuring", and the urgent establishment of what it called a "national forum on science".

The 'Science and Technology Initiative', as it has become known, was originally intended to act as such a forum. It was launched in May in response to a critical report on the country's research system sponsored by the African National Congress (ANC), the Congress of South African Trade Unions (COSATU), and the South African National Civic Organization (SANCO), and carried out by Canada's International Development Research Centre (IDRC).

While two subsequent plenary sessions have been held, the driving force behind the initiative has been an eight-member working group. But university scientists are rep-



South Africa's universities will need to adapt to a new environment.

resented on the working group only through the Committee of University Principals, and feel that this is inadequate. In particular, in calling for a moratorium on unilateral restructuring, the black scientists claim that what they see as a non-representative body should not be responsible for implementing changes in South African science.

One priority identified by the working group is the need to devise a new system for funding science in the new federal South Africa. The discussion on this topic is being co-ordinated by Roger Jardine, co-ordinator of science and technology planning for the ANC. He favours adoption of a model similar to the German Wissenschaftsrat, under which funds are provided by central government but administered by a council made up of scientists and political representatives from each region of the country.

To the surprise of some observers, the working group has chosen not to intervene in the process of allocating the science budget for the 1994/5 financial year. This has angered the universities, who feel that they are getting a raw deal under the current funding proposals (see *Nature* 356, 9; 1992).

As before, the Scientific Advisory Board has not made public its recommendations to the cabinet, a stance which the IDRC review criticized as resulting in a situation where "no independent assessment can be made of the extent, quality, relevance or impact of its advice".

Michael Cherry

Zoologist to head UK environment council

London. The latest piece in the new jigsaw of British science was put into place this week with the announcement that John Krebs will become chief executive of the Natural Environment Research Council from April. Krebs joins the line-up of scientists and industrialists who will head the newly structured research councils.

Krebs, a research professor in the department of zoology at the University of Oxford since 1988, takes control of the

new council at a time when its remit is being increased to include atmospheric chemistry, science-based archaeology and the UK government's favoured space project, Earth observation.

Widely known for his research on decision-making and memory in animals, Krebs is currently director of the Agriculture and Food Research Council's Unit of Ecology and Behaviour, and chairs the council's Animals Research Committee. □

Seismologists say Los Angeles quake sets 'ominous pattern'

Washington. This week's Los Angeles earthquake has reinforced seismologists' fears that California is on track for a seismic event far more potent than the one which registered 6.6 on the Richter scale at 4:31 a.m. on Monday morning.

As the death toll from the tremor mounted, leading seismologists were preparing to investigate how the event, epicentred in the San Fernando valley just over 30 kilometres north-west of central Los Angeles, would fit into two equally troubling patterns of seismic activity in the state of California.

One pattern, which has been observed worldwide, is a tendency for major shocks of this type to happen once a decade or so in the run-up to a catastrophic earthquake. Monday's event follows a shock of the same strength in the San Fernando valley in 1971.

Allan Lindh of the United States Geological Survey's Menlo Park research centre near San Francisco says that such a pattern preceded the great San Francisco earthquake of 1906, and has also been noted in Russia and Japan. Lindh believes the event to be "extremely ominous".

The second, more localized pattern is one of increasingly regular, low level tremors in southern California. Monday's event is the fifth to register more than five on the Richter scale in the area just north of Los Angeles since 1987, says Lucy Jones of the US Geological Survey's station at the California Institute of Technology at Pasadena.

Like previous tremors, it took place on one of a set of buried faults in the Earth's crust, where stress builds up due to the proximity of an S-bend on the San Andreas fault, which passes some 50 kilometres away.

The scientific follow-through from the earthquake will also try to refine estimates of the frequency of such events, find out which faults are active, and investigate just how big such secondary events — short of the dreaded primary event on the San Andreas — could be in the future.

Chris Scholz from the Lamont-Doherty Earth Observatory of Columbia University, New York, says the event does confirm that stresses are building up in California which will at some stage be released by a major earthquake, of magnitude 8 or more on the San Andreas.

Such concerns led the National Science Foundation to set up a Southern California Earthquake Center three years ago, at the University of Southern California. But the centre has so far only received about half of its expected \$5 million-a-year funding: after Monday that, at least, may change.

Colin Macilwain

AIDS debate continues

SIR — G. Garnett and R. Anderson (*Nature* 366, 716; 1993) make erroneous comments on our unpublished scientific data. Both ethics and etiquette demand that permission should be obtained by authors before citing unpublished work of others, but the authors did not do so on this occasion.

I emphasize that we are not involved in a research programme on AIDS in Africa, or elsewhere. Instead, we are concerned with a high-frequency cancer, Burkitt's lymphoma, of children in Africa; I have visited Malawi on several occasions to organize these investigations. Routinely, in order to ascertain how to handle safely in our London laboratory the biopsy materials received from Malawi, blood samples from the same children have been screened for two viruses, hepatitis-B virus (HBV) and HIV. Since 1987, around 300 such samples have been analysed.

Our findings, as mentioned informally to Anderson, are that whereas about 14 per cent of these are positive for the very infectious HBV, only about 1–2 per cent per year are positive for HIV — a figure unchanged over the 6-year period 1987–92. The number of patients with Burkitt's lymphoma seen over this same period in a single hospital in Malawi has increased by nearly 100 per cent. Patients range in age from 1 year to early teens, with a median age of about 6. Many, but by no means all, of these children come from rural areas.

Garnett and Anderson thus distort our data. Anderson believes our findings could be explained by the fact that HIV-positive children would have died before reaching the age when they might have developed Burkitt's lymphoma. This remains a possibility but is an unlikely explanation. I have offered the alternative explanation of a rural bias, with the caveat that if this were the case, then figures on HIV-positivity for entire countries such as Malawi could not, and should not, be deduced from urban data.

Garnett and Anderson further state that "Malawi is not a country from which data on the prevalence of HIV are readily available". This seems a misrepresentation. Their own review article (R. M. Anderson *et al.* *Nature* 352, 581–589; 1991; Table 1) cites HIV figures for Malawi as: weighted mean per cent seroprevalence, 17.0; number of surveys, 19; total number of people tested, 7,736; estimated urban population size, 1 million. (The total population of Malawi is nearly 10 million.) There are various other papers dealing with HIV in Malawi, including those of P. J. Miotti and colleagues (Johns Hopkins University School of Hygiene and Public Health) drawn from data derived from women in an antenatal clinic at a major urban hospital there, and a con-

trol case study (J. M. Pönnighaus and others) from the leprosy evaluation project in the rural northern part of Malawi.

Readers may ask why we have not yet published our own data. There are several explanations. One, we did not set out to study HIV, the presence of the virus being possibly relevant, but incidental, to our work on Burkitt's lymphoma. The presence of the virus is, however, relevant to safe sample handling. Our studies are continuing and not yet complete, and we have not sought permission from the local government to publish them. Further, and also important, many African countries resent the apparent deception practised on them with regard to HIV statistics (D. Serwadda and E. Katongole-Mbidde, *Lancet* 335, 842–843; 1990). We respect these sentiments, and would prefer to be on safe ground rather than rushing into print with apparently sensational, preliminary information.

Burkitt's lymphoma in children is at least as great a human tragedy as AIDS in Malawi or elsewhere in Africa, but there is very little funding for research on this topic. Are our priorities right?

Beverly Griffin

Department of Virology,
Royal Postgraduate Medical School,
London W12 0NN, UK

SIR — I welcome *Nature's* innovative approach to handling a systematic bias by a major London newspaper, as stated in the article "New-style abuse of press freedom" (366, 493; 1993). It is impossible to think of a disease that has been as politicized as AIDS by both conservatives and liberals, often in an attempt to advance a hidden or not so hidden agenda. It is therefore imperative to promote an accurate understanding of HIV infection and AIDS among scientific and medical professionals, as well as the lay public.

The article condemning the media bias of *The Sunday Times* has, perhaps inadvertently, fallen victim to another aspect of HIV infection subject to numerous political agendas: the reference to "those who consider the discomforts of 'safe sex' to be tiresome encumbrances" suggests great efficacy of the condom in decreasing HIV transmission.

The second-year (preclinical) medical students at the University of Arizona held a similar view of the condom, in that 89 per cent of the class believed the efficacy to be 98 per cent or higher. Ironically, some of them may well be the result of condom failures for the prevention of pregnancy. A recent meta-analysis of population-based studies of the efficacy of the condom in decreasing HIV transmission yielded an estimated efficacy of approximately 70 per cent¹, similar to the

50–90 per cent reduction in transmission noted for circumcision (reviewed in ref. 2). Interestingly, circumcision has seldom been touted for potential reduction of HIV transmission, even though it does not depend on patient compliance. Presumably this is because circumcision is less politically correct than condoms.

It is also interesting that a survey of the sexual practices of youth in Baltimore after the revelation of the HIV positivity of the basketball player Magic Johnson showed that condom use did not change, but the number of sex partners decreased³. It may be that the condomania of the early 1990s is only partly justified, and that a broad range of interventions should be considered in the effort to control the HIV pandemic.

Rodney D. Adam

University of Arizona Health Sciences
Center, Tucson, Arizona 85724, USA

1. Weller, S. C. *Soc. Sci. Med.* 36, 1635–1644 (1993).

2. Kreiss, J. K. & Hopkins, S. G. *J. inf. Dis.* 168, 1404–1408 (1993).

3. *MMWR* 42, 45–48 (1993).

SIR — I read your attack on *The Sunday Times* and the reply in that paper. Many people are worried, asking for the case to be properly studied. The unfalsifiability of the HIV hypothesis is at the core of the attack on it, yet there seems to be little genuine response to the objection.

The inclusion of positive HIV status within the diagnostic criteria of AIDS is a classic popperian *ad hoc* move. That inclusion is often implicit in the writings of the pro-HIV lobby but sometimes also explicit. Duesberg reasons that HIV must not be included as part of the diagnosis, and then, looking at the research, concludes that it is not a cause. That case now stands in a logically superior position to the HIV hypothesis. The worry is that to save the HIV hypothesis an absurd sleight of hand has been performed.

K. Ghattas

57 Harley Street, London W1N 1DD, UK

SIR — I am delighted to see you taking up the cudgels with such rigour and righteousness on behalf of responsible press reporting. With reference to the annotation under the title of your leading article, a British scientific journal, *Nature*, has so consistently misrepresented the issues of animal welfare, animal rights and wildlife conservation and protection that a group of nongovernmental organizations plan to monitor its future treatment of these issues, if only to save their members the trouble of buying it.

The 'device' you have offered for combating this problem is promising and can equally be adopted by those affected or offended by *Nature's* biased reporting.

Sidney J. Holt

Podere Il Falco,
06062 Città della Pieve (PG), Italy

The great multimedia revolution

Vice-President Al Gore's dream of an information superhighway has coincided in the United States with the ambitions of the media companies to provide a variety of services. One side-benefit may be cheaper supercomputers.

FOR much of the decade or so in which parallel computers (usually, and mystifying, called "massively parallel computers") have been with us, they have often seemed to be machines in search of useful functions. Anybody could quickly be convinced that a machine built by wiring together a large number of microprocessors (1,024, or 10^{10} , seems a favourite number) would be at least as computationally competent as the same number of computing elements working independently, and that the competence of the whole might be much greater than that of the sum of the parts, depending on how the wiring is done, or the "architecture" as they say in the trade.

At the outset, of course, as in the earliest days of electronic computing, the hardware was streets ahead of the software. Companies such as The Thinking Machine found it necessary to persuade bright sparks such as Stephen Wolfram (since then the originator of the program called Mathematica) to spend summer vacations in Boston wondering what to do with parallel computers, and how. With the fullness of time, the Pentagon has done its bit by buying the first machines; now there is a general understanding in the research community of what applications in science and technology will best suit parallel machines, and what will remain the chief outlets for vector machines such as those made by Cray Research and Fujitsu.

So who then would have thought that it would require the entertainment industry to create a substantial market for parallel computers? Even national lotteries cannot need computing power on such a scale and, even if they did, they would need only one machine each. A project to compile a database of all the chess games played, not simply by grand masters but by indifferent players as well, would be a great consumer of disk space but would require very little computation. What, then, can be the driving force behind the impending growth of the market for parallel computers? The most banal of all, as it happens: the ambition of several organizations to provide the world's television audiences with video films on demand, "VOD" as they say in that different trade.

A little reflection will show that parallel computers must be the answer. In the great nirvana of home entertainment, when people will be able to order up what they want to see, a person in front of a television screen who has an impulse to see some favourite film, say "High Noon", cannot be kept waiting after signalling the intention for fear that the impulse will dissipate, and with it the

revenue that would accrue. But does that mean that a remote provider of the video signal must keep several copies of every film that customers are likely to wish to see? And then maintain a small army of people to ensure that the appropriate film is fitted onto the output device connected with the customer who places the order? That is plainly a wasteful and error-prone way of dealing with the problem.

The use of parallel computers seems to be the response of Oracle Inc. of California, said to be the second software company (after Microsoft Inc.) in the world, to the ambition of telephone companies to send video signals corresponding to one of many films down telephone lines to each of many different customers. Plainly a substantial amount of image compression is necessary, especially if a simultaneous voice conversation is not to be impeded. The spare bandwidth in a twisted pair cannot be more than 1 kilohertz, there will be 1 megabyte of information (allowing for colour) in a still picture and there are 25 or 30 pictures a second in a video signal, so compression by a factor of 25,000 or more is necessary.

Although it is evidently more feasible to compress a series of related images (as in a film) than a sequence of discordant images, it will be interesting, in due course, to learn just how Oracle proposes to do the trick. But meanwhile, it is plain that the parallelism of the computing architecture is required not so much because of the complexity of that task, but so as to deal with the multiplicity of users and with the near (but not exact) simultaneity of their demands. Rarely can home entertainment have driven technology so powerfully — since the boom in computer games began ten years ago.

That may not be quite what US Vice-President Al Gore has had in mind the past several months when talking of the economic and technological benefits that will flow from the information "superhighway" he would have built in the United States, but there is no good reason why VOD should be scorned on that account. On the contrary, if there is to be a superhighway made of optical fibres, with at least one connection to every household, it is as well that there should be a business by means of which those who provide and maintain the cables can earn a return on their investment.

Whether Gore's vision of the superhighway will become reality and, if so, what its consequences will be are altogether different questions, but several US corporations evidently take the dream seriously.

Would two of the largest cable television companies in the United States be bidding up the price of Paramount, the Hollywood film-maker, to close on \$10 billion were it not that there is a shortage of material to pipe to peoples' television receivers? In the same few weeks, MCI (a long-distance and international communications company) has announced that it will spend an extra \$20 billion over the next six years on optical fibre networks in the United States. In a different field, Microsoft described last year its plan to hire more than 500 software engineers to staff a division aimed at generating multimedia products aimed at the computer in the living-room, not the study.

Gore is plainly doing all he can to help along this revolution. His talk on 11 January at Los Angeles confirmed what the US government has been saying for weeks — that there will be a further loosening of the regulations that at present restrict certain kinds of corporations to certain kinds of business. It is plainly a technological nonsense that both cable television and telephone companies should provide electronic connections to individual houses, but should be forbidden to take in each other's washing. Providers of database services, meanwhile, are restricted technically by the capacity of telephone lines and commercially by the individual ownership of modems, the costs of what they provide are too high to allow for the free dissemination of information, even knowledge.

Outside the United States, the regulations and the restrictions are still more onerous. Britain, where the cable companies are being encouraged to offer telephone services to make up for the decision to privatize the nationalized telecommunications company (BT) as a single entity, is an exception. (BT is hoping to fight back with VOD by using Oracle's software.) But elsewhere in Europe, the constitutional flexibility required for the equivalent of Gore's superhighway is itself still a dream.

Does that matter? The true benefits of Gore's superhighway are still a long way off, but if the concept can be made to work, it could do everything he claims for it. The benefits of wide digital access to educational tools could, for example, be the shot in the arm that public education has for decades been crying for. The benefits (for many) of not having to travel to work are incalculable. And if VOD drives down the price of parallel computers, that could also be a benefit. The superhighway is not a fantasy.

John Maddox

From the outside in

Robert C. Haddon

In their original paper on buckminsterfullerene, Kroto *et al.*¹ suggested that the NMR of the central atom should show remarkable chemical shifts because of the ring currents induced in the delocalized π -electron ring system of C_{60} . This point has been much discussed and disputed, but on page 256 of this issue Saunders *et al.*² provide striking evidence for the statement. In this instance, though, it is C_{70} that shines most brightly. By pressurizing ^3He in the presence of a C_{60}/C_{70} mixture, Saunders *et al.* are able to obtain endohedral fullerenes suitable for solution characterization by NMR spectroscopy. They show that the ^3He nuclei inside C_{60} and C_{70} are magnetically shielded by 6 and 29 p.p.m., respectively, with respect to free ^3He nuclei.

At first sight, Kroto and co-workers' original suggestion seems eminently reasonable. If C_{60} is taken as the three-dimensional analogue of two-dimensional benzene, then the π -electrons which are responsible for the enhanced diamagnetism and highly anisotropic NMR shielding of the benzene molecule should function in the same manner in C_{60} , except that they should now escape their two-dimensional confines and instead be free to circulate on the surface of a sphere in response to an applied magnetic field. In short, this would be the perfect molecular analogue of the Larmor precession experienced by atomic electrons. This latter phenomenon formed the basis for the Pauling free-electron theory of molecular ring currents. Indeed when the free-electron theory is applied to the C_{60} molecule, a diamagnetic susceptibility 41 times that of the ring-current magnetic susceptibility of benzene is calculated (assuming the electronic circulations to occur at a radius of 3.47 Å; ref. 3). Such currents would produce a large chemical shift (shielding) of $\delta_{\text{RC}} \approx -100$ p.p.m. at the centre of C_{60} .

Cancellation

The London theory is the quantum mechanical counterpart of the classical Pauling theory. In many cases the two methods give rise to similar results, but in the case of C_{60} , the London theory predicts a vanishingly small ring-current magnetic susceptibility³. Because of its classical origins the Pauling theory treats all electronic circulations as diamagnetic, but the London theory correctly includes the possibility of paramagnetic ring currents (from the van Vleck term) and the existence of such circulations — usually in antiaromatic molecules — has received strong experimental support. So when the fullerenes were isolated and their magnet-

ism examined experimentally^{4,5}, the very small magnetic susceptibility was interpreted in terms of a cancellation of the diamagnetism by a very large van Vleck paramagnetism⁵, in agreement with the predictions of the London theory.

But the nature of the ring currents in the fullerenes remained a mystery until 1992, because the topological complexity of the molecules had precluded their theoretical evaluation. Finite-field solution of the London equations evaluates the magnetic susceptibility but not the underlying ring currents³. As the effect of ring currents on the carbon atoms themselves is somewhat uncertain, experimental approaches to the problem were hampered by the lack of protons in the fullerenes (proton NMR spectroscopy is the usual probe because the chemical shift provides a measure of the secondary magnetic field arising from the ring currents).

In 1992, methods were developed to estimate the ring currents within the London theory⁶ and chemists managed to synthesize fullerene derivatives in which the electronic structure was largely intact (fulleroids)⁷, and which positioned protons close to the (outer) surface of the fullerenes^{8–11}. Although only a handful of compounds have been examined so far, both theory and experiment suggest that there are strong ring currents in C_{60} , and these currents move the chemical shifts of the nearby nuclei by a few parts per million. Remarkably, the five-membered rings show paramagnetic ring currents whereas the six-membered rings show diamagnetic ring currents, and these cancel almost entirely in their contribution to the magnetic susceptibility⁶. So the magnetic fields outside the fullerene surface produce chemical shifts in the attached protons^{6–10} quite characteristic of those previously observed in bridged annulenes¹².

The result obtained by Saunders and colleagues makes it difficult to turn this picture outside in. After all, the magnetic susceptibility and the exterior shielding contours are fairly well established and the enhanced diamagnetic ring current found in C_{70} (refs 4, 5) is apparently reflected in the endohedral chemical shift, which is much larger for helium trapped in C_{70} than in C_{60} (ref. 2). The usual way of applying the London theory of chemical shifts makes use of line currents to approximate the electronic circulations, and this is only valid when the test nuclei are far from the current flow⁶. It is possible that the interior lobes of the rehybridized π -orbitals produce electronic circulations that are different from those on the exterior and are not captured by this

simple (line) current distribution. If so, a more sophisticated treatment will be needed to account for the current flow on the interior of the fullerenes¹³.

Nevertheless, the chemical shift measured for $^3\text{He}@C_{60}$ by Saunders *et al.* can be accommodated without complications given a diamagnetic C_{60} ring current of about 2 to 3 times that of benzene. This is just the value expected from the London theory when the two different C_{60} bond lengths are taken into account^{3,4}. Even without this refinement, the observed chemical shift of $^3\text{He}@C_{70}$ is in good agreement with the London calculations⁶.

Characterization

It will be interesting to see the results of this technique in the case of C_{60}^{6-} , where theory^{3,6} already predicts a very large ring current diamagnetism and the ^{13}C NMR spectra of the reduced fullerenes show intriguing chemical shifts¹⁴. The higher fullerenes, fullerene derivatives and superconducting fullerenes are all inviting subjects for ^3He NMR study. Given the large range of chemical shifts already demonstrated by Saunders *et al.*, this technique provides a sensitive probe of the molecular and electronic structure of fullerenes. If it proves possible routinely to inflate the fullerenes with ^3He , this experiment would be a particularly useful analytical technique in a field that suffers from a lack of characterization methods.

These results and previous reports by other authors all point to the importance of ring currents in the fullerenes, implying that the π -electrons are indeed able to circulate in the presence of a magnetic field. Although the details of the circulations seem to be unique to the fullerenes, we are coming closer to understanding the electronic structure of these compounds. In particular it is now clear that they must be interpreted as delocalized π -electron systems. Kroto and colleagues¹ were right to assert that the C_{60} molecule "appears to be aromatic". □

Robert C. Haddon is at AT&T Bell Laboratories, 600 Mountain Avenue, P.O. Box 636, Murray Hill, New Jersey 07974, USA.

- Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–163 (1985).
- Saunders, M. *et al.* *Nature* **367**, 256–258 (1994).
- Elser, V. & Haddon, R. C. *Nature* **325**, 792–894 (1987); *Phys. Rev. A* **36**, 4579–4584 (1987).
- Haddon, R. C. *et al.* *Nature* **350**, 46–47 (1991).
- Ruoff, R. S. *et al.* *J. phys. Chem.* **95**, 3457–3459 (1991).
- Pasquarello, A. *et al.* *Science* **257**, 1660–1661 (1992); *Phys. Rev. A* **47**, 1783–1789 (1993).
- Wudl, F. *Accs chem. Res.* **25**, 157–161 (1992).
- Prato, M. *et al.* *J. Am. chem. Soc.* **115**, 7876–7877 (1993).
- Prato, M. *et al.* *J. Am. chem. Soc.* **115**, 8479–8480 (1993).
- Isaacs, L., Wehrsig, A. & Diederich, F. *Helv. chim. Acta* **76**, 1231–1250 (1993).
- Smith, A. B. III *et al.* *Proc. Mat. Res. Soc. Mtg* 1993.
- Garratt, P. J. *Aromaticity* (Wiley, New York, 1986).
- Fowler, P. W., *et al.* *Chem. Phys. Lett.* **179**, 174–180 (1991).
- Bausch, J. W. *et al.* *J. Am. chem. Soc.* **113**, 3205–3206 (1991).

Towards a better aspirin

John Vane

In 1971, I proposed that the mechanism of action of the aspirin-like drugs was through their inhibition of prostaglandin biosynthesis¹. Since then, there has been intense interest in the interactions between this diverse group of inhibitors and the enzyme known to pharmacologists as cyclooxygenase (COX). It exists in two isoforms, COX-1 and COX-2, and several new drugs have reached the market based on COX-1 enzyme screens.

Picot, Loll and Garavito (page 243 of this issue²) have now given fresh impetus to the field by determining the three-dimensional structure of COX-1 (PGHS-1, for prostaglandin H₂ synthase-1, in their terminology). The structure, although at low resolution, is undoubtedly correct and indicates that the enzyme is in a monotopic arrangement with respect to the cell membrane (that is, it integrates into only a single leaflet of the lipid bilayer). This is the first three-dimensional structure of such a protein, and the stereo pictures in the paper are fascinating; those who, like myself, cannot bring the images into three dimensions by squinting at them, should equip themselves with a pair of stereo glasses to do them full justice.

COX first oxidizes arachidonic acid to prostaglandin G₂ through a cyclooxygenase activity and then peroxidizes prostaglandin G₂ to prostaglandin H₂. It is thus a bifunctional enzyme, and comprises three independent folding units: an epidermal growth factor-like domain, a membrane-binding motif and an enzymatic domain. The sites for peroxidase and cyclooxygenase activity are adjacent but spatially distinct. Three of the helices of the structure form the entrance to the cyclooxygenase channel and their insertion into the membrane could allow arachidonic acid to gain access to the active site from the interior of the bilayer.

The cyclooxygenase active site is a long, hydrophobic channel, and Garavito and colleagues present arguments that some of the aspirin-like drugs such as flurbiprofen inhibit COX-1 by excluding arachidonate from the upper portion of the channel. Tyrosine 385 and Ser 530 are at the apex of the long active site. Aspirin irreversibly inhibits COX-1 by acetylation of Ser 530, thereby excluding access for arachidonic acid³. The S(-) stereoisomer of flurbiprofen interacts, through its carboxylate, with Arg 120, thereby placing the second phenyl ring within van der Waals contact of Tyr 385. The authors suggest that there are a number of other sub-sites for drug binding in the narrow channel.

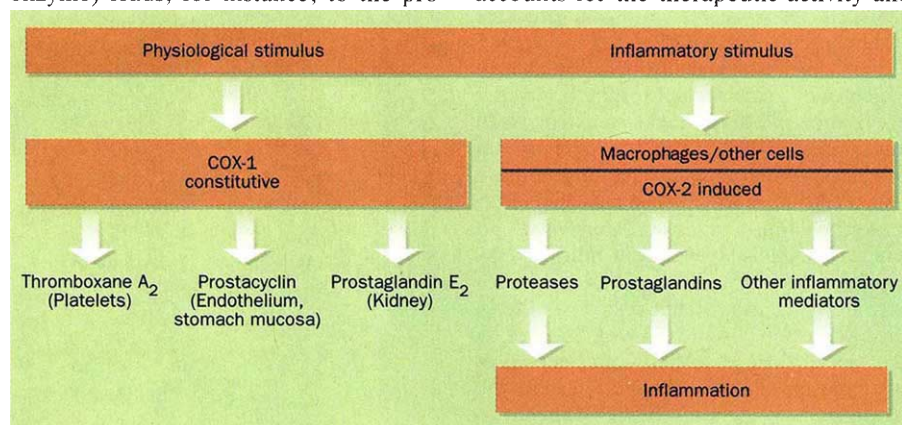
A few years ago, the publication of the three-dimensional structure of COX-1

would have caused enormous interest in the drug industry as a pointer to finding new and better aspirin-like drugs. However, hundreds of medicinal chemists around the world are trying to find non-steroid anti-inflammatory drugs (NSAIDs) which do *not* interfere with COX-1. This is because attention is highly focused on selective inhibitors of an inducible COX-2.

Activation of COX-1 (the constitutive enzyme) leads, for instance, to the pro-

enzymes⁷. Aspirin, indomethacin and ibuprofen are much less active against COX-2 than COX-1 (ref. 8). Indeed, the two strongest inhibitors of COX-1 are aspirin and indomethacin, the two NSAIDs which cause the most damage to the stomach⁹. The spectrum of activities of some ten standard NSAIDs against the two enzymes ranges from a high selectivity towards COX-1 (150-fold for aspirin) through to equi-activity on both¹⁰. This range of activities nicely explains the variations in the side-effects of NSAIDs at their anti-inflammatory doses.

Over the years, the theory that inhibition of prostaglandin formation accounts for the therapeutic activity and



Actions of the two known isoforms of COX (PGHS). The hypothesis stemming from this scheme is that the therapeutic effects of drugs such as aspirin are due to inhibition of COX-2, whereas the unwanted side-effects (and the action on platelets) stem from inhibition of COX-1.

duction of prostacyclin, which when released by the endothelium is anti-thrombogenic⁴ and by the gastric mucosa is cytoprotective⁵. These are clearly physiological functions for COX-1. The second isoform, COX-2, is induced in a number of cells by pro-inflammatory stimuli⁶. Luckily for the medicinal chemists, the two isoforms show only 60 per cent homology of both their nucleic-acid and amino-acid structures: while the messenger RNA encoding COX-2 is 4.1 kilobases in length, that of COX-1 and its homologues from different species is only 2.8–3.0 kilobases. The discovery of the inducible isoform⁶ allows a re-interpretation and refinement of the general theory that inhibition of COX activity explains the therapeutic and side-effects of the aspirin-like drugs.

Because COX-2 is induced in migratory and other cells by inflammatory stimuli and by cytokines, it is attractive to suggest that the anti-inflammatory actions of aspirin and its fellows are due to the inhibition of COX-2, whereas the unwanted side-effects (such as irritation of the stomach lining and toxic effects on the kidney) are due to inhibition of the constitutive enzyme, COX-1 (see figure). The importance of the discovery of the inducible isoform is highlighted by the differences in pharmacology of the two

the side-effects of the aspirin-like drugs has been challenged, notably by Weissman¹¹. His arguments were partly based on comparing the actions of salicylate and aspirin, which are said to be equally effective against arthritis in the clinic¹², whereas in the original observations¹ salicylate was much weaker as an inhibitor of COX than aspirin. This may now well be explained by the existence of the two isoforms, for both drugs are comparably weak inhibitors of COX-2 in whole-cell preparations⁷, although the mechanism of action of salicylate may be compounded by a suppression of the induction of COX (ref. 13).

Paracetamol also posed a problem for the original theory, for in therapeutic doses it has weak anti-inflammatory activity but is a stronger analgesic and antipyretic¹⁴. In 1972, Flower and I showed that COX preparations from the brain were more sensitive to paracetamol than those from the spleen and suggested that there may be different COX isoforms¹⁵. In the light of recent discoveries, perhaps there is also a COX-3 on which paracetamol has a preferential action.

All the results so far published (and many yet to be published from the drug industry), support the hypothesis that the unwanted side-effects of NSAIDs are due

to their ability to inhibit COX-1 whereas their anti-inflammatory (therapeutic) effects are due to inhibition of COX-2. Other roles for COX-2 will surely be found in the next few years, for prostaglandin formation is under strong control in organs such as the uterus. The identification of selective inhibitors of each isoform will not only provide an opportunity to test the new hypothesis but also lead to advances in the therapy of inflammation. As selective inhibitors of COX-2 come to the market, so the present clutch of NSAIDs will become obsolete. Let us hope that Garavito and his colleagues, who have made such a striking contribution by mapping the three-dimensional structure of COX-1, will soon produce similar pictures of COX-2.

However, there is one use for aspirin which will expand, and that is in the prevention of heart attacks and strokes. In low doses it has a unique activity on platelets, which almost certainly contain COX-1. It is unique for two reasons. First, platelets cannot regenerate enzymes, so when COX-1 is irreversibly inhibited by aspirin, the platelet cannot produce thromboxane A₂ for the rest of its life in the circulation (up to 10 days). Second, as the aspirin is absorbed into the pre-systemic circulation, the platelets in the blood meet it in much higher concentrations than in the arterial circulation. Thus, low-dose aspirin (75 mg rather than the usual tablet of 325 mg) has a selective effect on COX-1 in platelets, sparing that in the endothelial cell. By reducing thromboxane A₂ formation in platelets, aspirin reduces the risks of heart attacks and strokes by up to 50 per cent (ref. 16). Furthermore, it is a cheap drug and as such is unlikely to be bettered by new inhibitors of COX-1. □

John Vane is at the William Harvey Research Institute, St Bartholomew's Hospital Medical College, Charterhouse Square, London EC1M 6BQ, UK.

OSTEOPOROSIS

Boning up on genes

Gregory R. Mundy

It often raises eyebrows when it is suggested that common and apparently multifactorial polygenic diseases are linked to a single, critical gene. The latest example is osteoporosis. On page 284 of this issue¹, Morrison *et al.* report that bone mass, one of the main determinants of osteoporotic fracture, has a genetic component which can be ascribed largely to a simple allelic change in the receptor for 1,25 dihydroxy-

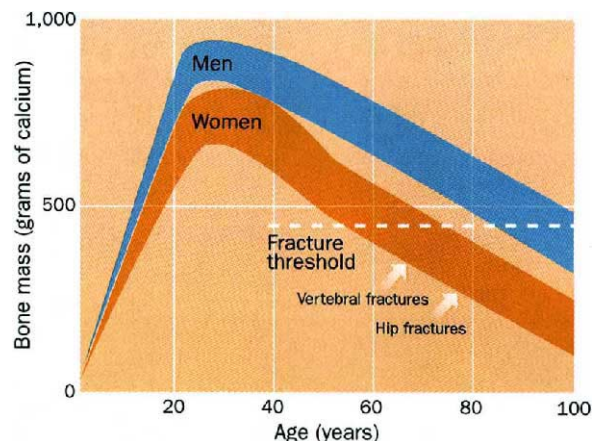
years, techniques that assess bone mineral density at different sites in the skeleton have indicated that monozygotic twins show less variation in bone mass than dizygotic twins³⁻⁵. (The non-genetic factors that influence bone mass, and their relationship to osteoporotic fracture, are indicated in the figure caption). What is new about the study of Morrison *et al.* is the conclusion that the main mechanism

responsible for this genetic effect (accounting for up to 75 per cent at important skeletal sites) is linked to polymorphism in the vitamin D receptor gene. Concordant dizygotic twins showed similar variations to monozygotic twins, further emphasizing the importance of this allelic change to bone mass.

At first glance, this finding may seem unlikely. Bone mass at any point in adult life is determined by complex interactions between bone resorbing (osteoclastic) and bone forming (osteoblastic) cells, whose integrated activities are controlled by a multitude of systemic hormones, local cytokines and growth regulatory factors. There may be literally hundreds of genes that could determine bone mass. But if one had to

predict the particular genes that are probably preeminent, the gene for the vitamin D receptor would surely rank highly. 1,25 dihydroxyvitamin D₃, the hormonal form of the vitamin, is required for the normal mineralization of bone, absorption of calcium from the gut, control of calcium and phosphate homeostasis, and regulation of parathyroid hormone secretion. Because the vitamin D receptor is ubiquitously distributed, there may well be actions of this hormone that are yet to be discovered. Moreover, the vitamin D receptor could also be acting through apparently non-genetic factors, such as dietary calcium intake which is a determinant of peak bone mass⁶. The vitamin D receptor is required for normal calcium absorption from the gut.

Although the results of Morrison *et al.* are exciting, they should be regarded cautiously until more work is done. The study was performed in a rather restricted population of Caucasians of mainly Anglo-Irish descent in Australia, so first, and most obviously, the population base needs to be extended to ethnic groups with



Changes that occur in bone mass with age. Factors involved in determining peak bone mass are genetic and non-genetic (nutrition, smoking, exercise, hypogonadism); factors increasing rate of bone loss in later life are ageing, the menopause and various lifestyle factors (including drug and alcohol intake). The risk of fracture due to osteoporosis occurs when bone mass falls below a theoretical fracture threshold. The results of Morrison *et al.*¹ suggest that the major genetic component responsible for bone mass is linked to polymorphism in the gene for the vitamin D receptor.

vitamin D, one of the hormones controlling calcium metabolism.

Osteoporotic fracture is a catastrophic event in an increasingly ageing population, so this finding will attract a lot of attention. Osteoporosis affects one in four postmenopausal white women, but also occurs frequently in men and women of all ethnic groups. Fracture of the hip (neck of femur), the most significant complication, afflicts some 1.7 million people worldwide each year, a number which is expected to grow to over 6 million a year by 2050 (ref. 2). A quarter of these patients die within 6 months as a direct consequence of the fracture and many of the survivors will require permanent, institutionalized care. Low bone mass is one of the principal risk factors for hip fracture (as well as other osteoporotic fractures), and determining the factors that optimize bone mass, and removing or reversing those which decrease it, are major goals of research into treatments for osteoporosis.

Few will be surprised that we are born with genes that determine our bone mass (or bone mineral density). For over 20

1. Vane, J. R. *Nature* **231**, 232-235 (1971).
2. Picot, D., Loll, P. J. & Garavito, R. M. *Nature* **367**, 243-249 (1994).
3. Roth, G. J., Stanford, N. & Majerus, P. W. *Proc. natn. Acad. Sci. U.S.A.* **72**, 3073-3076 (1975).
4. Moncada, S., Gryglewski, R., Bunting, S. & Vane, J. R. *Nature* **327**, 663-665 (1976).
5. Whittle, B. J. R., Higgs, G. A., Eakins, K. E., Moncada, S. & Vane, J. R. *Nature* **284**, 271-273 (1980).
6. Xie, W., Robertson, D. L. & Simmons, D. L. *Drug Deliv. Res.* **25**, 249-265 (1992).
7. Mitchell, J. A., Akaraseenont, P., Thiemeermann, C., Flower, R. J. & Vane, J. R. *Proc. natn. Acad. Sci. U.S.A.* **90**, 11693-11697 (1993).
8. Meade, E. A., Smith, W. L. & DeWitt, D. L. *J. biol. Chem.* **268**, 6610-6614 (1993).
9. Lanza, F. L. *Scand. J. Gastroenterol.* **24**, 24-31 (1989).
10. Akaraseenont, P., Mitchell, J. A., Thiemeermann, C. & Vane, J. R. *Br. J. Pharmacol.* (in the press).
11. Weissman, G. J. *Lipid Med.* **6**, 275-286 (1993).
12. Weissman, G. *Scient. Am.* **84**-90 (January, 1991).
13. Wu, K. K., Sanduja, R., Tsai, A. L., Ferhanoglu, B. & Loose-Mitchell, D. S. *Proc. natn. Acad. Sci. U.S.A.* **88**, 2384-2387 (1991).
14. Clissold, S. P. *Drugs* **32** (Suppl. 4), 46-59 (1986).
15. Flower, R. J. & Vane, J. R. *Nature* **240**, 410-411 (1972).
16. Vane, J. R. & Botting, R. M. (eds) *Aspirin and Other Salicylates* (Chapman & Hall, London, 1992).

different bone mass, such as African-Americans, Mexican-Americans and Orientals. The allelic change in the vitamin D receptor is simple and unambiguous, and should be relatively easy to examine in other population groups. Second, more studies are needed to show if this polymorphism results in alterations of receptor action, a necessary prerequisite for a direct functional link between the observed polymorphism in the gene and peak bone mass. A preliminary experiment described in the paper¹ shows an increase in luciferase activity when this reporter gene is linked to the 3' untranslated region of the vitamin D receptor, suggesting it is directly responsible for the functional change. Even more helpful would be the demonstration that administration of vitamin D to individuals with unfavourable alleles has a demonstrable difference in response parameters compared with similar treatment in individuals with different allelic patterns.

If these associations of bone mass with the vitamin D receptor gene hold up, then the implications are considerable. The place for biologically active forms of vitamin D in treatment of osteoporosis is controversial, some studies indicating no beneficial effect⁷, but others pointing to an increase in bone mineral density and reduction in vertebral fractures in patients with established disease^{8,9}, particularly when used in larger (and potentially more toxic) doses. Treatment with small doses of vitamin D and calcium has recently been shown to decrease rates of hip fracture in elderly patients¹⁰. If the vitamin D receptor gene determines bone mass, there might be little point in beginning treatment with vitamin D metabolites after bone mass has irreversibly declined, because of the effects of multiple non-genetic factors. On the other hand, 1,25 dihydroxyvitamin D treatment may be most effective in optimizing bone mass when used earlier in life in patients with unfavourable vitamin D receptor alleles.

The implications for research into the physiological effects of vitamin D on bone may be equally interesting. The active metabolites of vitamin D have myriad effects *in vitro* on cells in both the osteoclast and osteoblast lineages, as well as on other cells in the bone microenviron-

ment including lymphocytes, monocytes and macrophages. There are many potential mechanisms by which the vitamin D receptor could influence the delicate relationship between osteoclasts and osteoblasts which bone biologists will need to unravel. The vitamin D receptor is a transcription factor which forms homodimers or heterodimers with other members of the steroid hormone receptor superfamily (most notably, the retinoic acid receptor RXR)¹¹, so several signalling pathways may mediate effects on gene expression.

INERTIAL FUSION

Rayleigh's challenge endures

Richard D. Petrasso

AMIDST the news of the results¹ from the Princeton tokamak fusion reactor, we should not forget the other runner in the race to achieve practical fusion power. At an international conference on plasma physics last year*, the consensus was that an instability first recognized by Lord Rayleigh last century remains one of the most challenging problems in the field of inertial confinement fusion (ICF).

Lord Rayleigh, his interest sparked by experiments on cirrus cloud formation, derived in 1880 the properties of an interface instability that occurs when a dense fluid is supported by a lighter one². Invert a glass of water and this instability occurs — the dense water falls out even though, in a carefully prepared experiment, the air pressure is more than sufficient to support the water³. As the effects of gravity are, by the equivalence principle of general relativity, tantamount to acceleration in the opposite direction, the instability will also occur whenever a light fluid pushes against a dense one, as, for example, in supernovae explosions⁴, magnetically confined plasmas⁵ and heavy-nuclei collisions⁶. In the spectacular instance of supernova SN1987A, the unexpectedly early appearance of energetic γ -rays, following the rebound of the collapsed stellar core, was explained by long, dense fingers of radioactive ⁵⁶Co piercing the turbid blanket of tenuous gas that enveloped it and retarded its outward expansion⁴.

At the very least, this report of potential functional allelic changes in the vitamin D receptor gene is highly provocative. Whether or not it is confirmed, it will no doubt prompt other studies on the mechanisms responsible for the unquestionably important genetic influence on bone mass. □

Gregory R. Mundy is in the Division of Endocrinology and Metabolism, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, Texas 78284, USA.

Plasma physicists grappling with the process of inertial confinement fusion have focused their attention on this interface instability for the past decade, and are now making substantial progress. In the ICF approach, a small quantity of fusion fuel (consisting of the hydrogen

IMAGE
UNAVAILABLE
FOR COPYRIGHT
REASONS

Lord Rayleigh — legacy of instability.

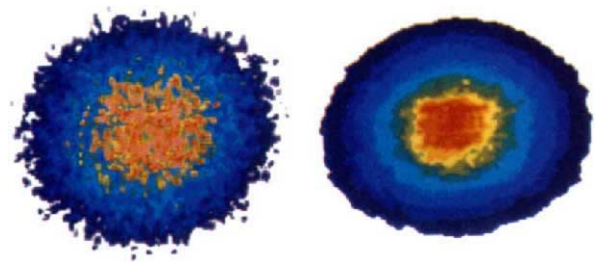
isotopes of deuterium and tritium) is to be imploded and compressed in about 10^{-8} s by ablating (that is, explosively evaporating) a thin, dense, spherical shell that surrounds the fuel. Before compression, the sphere radius and the shell thickness are about 2 mm and 0.2 mm, respectively. At the moment of peak compression, when the sphere radius is smaller by a factor of 20 and the fuel density is 1,000 times that of solid deuterium, the deuterium-tritium fuel, which is heated by compressional work, ignites and releases around 500 MJ of fusion energy (equivalent to 3.5 gallons of oil). But the

1. Morrison, N. A. *et al.* *Nature* **367**, 284–287 (1994).
2. Melton, L. J. *Bone* **14**, S1–S8 (1993).
3. Smith, D. M. *et al.* *J. Clin. Invest.* **52**, 2800–2808 (1973).
4. Pocock, N. A. *et al.* *J. Clin. Invest.* **80**, 706–710 (1987).
5. Dequeker, J. *et al.* *Bone* **8**, 207–209 (1987).
6. Johnston, C. C. *et al.* *New Engl. J. Med.* **327**, 82–87 (1992).
7. Ott, S. M. & Chesnut, C. H. *Ann. intern. Med.* **110**, 267–274 (1989).
8. Tilyard, M. W., Spears, G. F. S., Thomson, J. & Dovey, S. *New Engl. J. Med.* **326**, 357–362 (1992).
9. Gallagher, J. C. & Goldgar, D. *Ann. intern. Med.* **113**, 649–655 (1990).
10. Chapuy, M. C. *et al.* *New Engl. J. Med.* **327**, 1637–1642 (1992).
11. Carlberg, C. *et al.* *Nature* **361**, 657–660 (1993).

* American Physical Society, Division of Plasma Physics Annual Meeting, St Louis, 1–5 November 1993.

spherical ablating surface is unquestionably vulnerable to this interface instability, and the most urgent issue is whether sufficient fuel compression will occur before the instability dismantles the ablating shell, mixing it with the heated fuel^{7,8}.

During the Manhattan Project in the Second World War, Geoffrey Taylor looked at the analogous consequences of this interface instability for the implosion dynamics of the atomic bomb (refs 9, 10; and P. Morrison, personal communica-



The effects of beam smoothing are dramatically evident in these cross-sectional images of a laser beam: left, without smoothing; right, with smoothing (J. Knauer, University of Rochester). Without smoothing, the laser beam can seed and then drive the Rayleigh–Taylor instability, which, for fusion capsules, would thwart adequate fuel compression.

tion). Thereafter, the instability was referred to as the Rayleigh–Taylor (RT). With the colossal differences in scale and design, and the use of highly reactive fissile material (²³⁵U and ²³⁹Pu), however, the instability is much less of a problem for an atomic weapon than for ICF.

Our present understanding is that if the instability grows unabated at the Rayleigh rate, then adequate fuel compression cannot be achieved and the ICF concept will fail (J. Kilkenny, Lawrence Livermore National Laboratory (LLNL)). Rayleigh's rate, γ_R , is $(2\pi g/\lambda)^{0.5}$, where g is the acceleration at the ablation front, and λ is the wavelength of the perturbation. Fortunately, the dynamics of plasma flowing outward across the ablation front reduce the instability growth rate from Rayleigh's value^{11,12}. The consequences of this ablative stabilization, plus other secondary ameliorations, are often expressed as a reduced growth rate $\gamma = \epsilon\gamma_R$, where ϵ is less than 1 and is a function of λ (K. Nishihara, Osaka University). The most unstable λ is of the order of the shell thickness.

In the linear regime, the size of the instability is well described by $a_0 \exp(\epsilon\gamma_R t)$, where a_0 is the amplitude of the initial perturbation, and t is time. Because of exponential growth (a valid model until the instability size is of order $\lambda/10$), even a small reduction in the growth rate can have dramatic consequences. Efforts to control the RT instability have in part focused on making surface imperfections as small as possible ($a_0 \sim 250$ Å) so that meaningful measurements of ϵ can be inferred during the linear phase. In fact by

the time $\epsilon\gamma_R t \sim 6$, three-dimensional calculations show that nonlinear effects generate and couple together different modes (J. Dahlburg, Naval Research Laboratory). This coupling between λ and its harmonics coincides with the dismantling of the ablation shell.

Parallel advances in laser-beam smoothing techniques^{8,13–15} are reducing nonuniformities in the laser irradiance which would otherwise create large initial perturbations from which the RT instability would spring. (In addition, these laser nonuniformities drive a different class of undesirable instabilities known as laser–plasma or parametric instabilities; R. Berger, LLNL). Carefully controlled experiments are in progress in which various levels of laser-beam smoothing are used on one arm of the NOVA glass laser (S. Glendinning, LLNL). For the KrF gas laser at the Naval Research Laboratory, root-mean-square nonuniformities in a single laser beam are about 1.6 per

cent (T. Lehecka, Science Applications International Corporation). When the system is fully operational with all 44 independent, overlapping beams, r.m.s. nonuniformity of about 0.25 per cent ($1.6/\sqrt{44}$) is expected, which is comparable to surface imperfections of ~ 200 Å. Another promising line is to generate, through an X-ray flash, a uniform plasma around the ablator, which should further reduce laser imprinting (R. S. Craxton, University of Rochester; O. Willi, Imperial College London).

With high levels of laser smoothing on NOVA, the effects of deliberately introduced initial perturbations can be seen (J. Knauer, University of Rochester). A low-Z (low mass number) plastic target was prepared with a perturbation of 30 μ m wavelength (λ) and 0.25 μ m amplitude (a_0). On irradiation, ϵ of about 0.65 was achieved when the hydrogen/carbon plastic was lightly doped with 6 per cent of chlorine. The desirable effects of the chlorine were manifested by an increase in the ablation rate and, secondarily, by an increase in the density scale length at the ablation front (according to theory, this increase should be beneficial; R. Betti, University of Rochester). But this doping procedure involves a delicate balance: premature heating of the fuel from the chlorine X-rays means that more laser energy is required to compress the fuel. This degrades the figure-of-merit known as the fusion gain (G), which is the ratio of fusion to laser energy, and which must be at least 30 for just-plausible designs of future ICF reactors (C. Verdon, University of Rochester).

In a different approach to achieving compression while minimizing the RT instability, either lasers (A. Hauer, Los Alamos National Laboratory) or particle beams (J. Quintenz, Sandia National Laboratory) can be used to generate X-rays that flood the interior of a cavity called a hohlraum^{7,8}. In turn these X-rays irradiate the ablator. (In the lexicon of the field, this is known as 'indirect drive', in contrast to 'direct drive' in which lasers directly irradiate the ablator.) In the X-ray drive experiments at LLNL, values of $\epsilon \approx 0.5$ were achieved for wavelength perturbations of around 50 μ m (Kilkenny; B. Remington, LLNL). These low values of ϵ result from the deep penetration of the X-rays into the ablator, increasing the plasma ablation rate.

Encouraging progress has been made in ameliorating the RT instability during the acceleration phase of the compression. But it is a herculean task: for like Hydra, the instability rears an ugly head elsewhere, this time just inside the sphere at the fuel–shell interface. During the final deceleration stage to peak compression, the less dense deuterium–tritium fuel now pushes vigorously against the denser shell, which is still inwardly imploding. Just as in supernova SN1987A, these conditions are ripe for the RT instability. But whereas in the acceleration phase the stabilizing effects of ablation reduced the growth to less than the Rayleigh value, now the full force of the instability takes hold. Still, there is reason for optimism as its onset at the fuel–shell interface commences only in the very last moments of compression. The vexing question is, will the shell remain sufficiently intact or will spikes of it penetrate and mix into the fuel, diluting its potency and subverting the fusion gain? In the near future, an upgraded laser facility (T. Boehly, University of Rochester) and new experiments (O. Landen, LLNL) will test this next interface for the effects of the irrepressible Rayleigh–Taylor instability. □

Richard D. Petrasso is in the Plasma Fusion Center, Massachusetts Institute of Technology, 175 Albany Street, Cambridge, Massachusetts 02139, USA.

1. MacIlwain, C. *Nature* **366**, 600 (1993).
2. Lord Rayleigh *Proc. Lond. math. Soc.* **14**, 170–177 (1883).
3. Lewis, D. J. *Proc. R. Soc. A* **A202**, 81–96 (1950).
4. Fryxell, B. *et al. Astr. J.* **367**, 619–626 (1991).
5. Curzon, F. L. *et al. Proc. R. Soc. A* **A257**, 386 (1960).
6. Moretto, L. S. *et al. Phys. Rev. Lett.* **69**, 1884–1887 (1992).
7. Craxton, R. S., McCrory, R. L. & Souders, J. M. *Scient. Am.* **255**, 68–79 (1986).
8. Lindl, J. D., McCrory, R. L. & Campbell, E. M. *Phys. Today* **45**, 32 (1992).
9. Hawkins, D. *et al. Manhattan District History, Project Y, The Los Alamos Project, LAMS-2532*, Vol. I (1961).
10. Taylor, G. I. *Proc. R. Soc. A* **A201**, 192–196 (1950).
11. Bodner, S. E. *Phys. Rev. Lett.* **33**, 761 (1974).
12. Takabe, H. *et al. Phys. Fluids* **28**, 3676 (1985).
13. Kato, Y. *et al. Phys. Rev. Lett.* **53**, 1057 (1984).
14. Lehmberg, R., Schmitt, A. & Bodner, S. J. *appl. Phys.* **62**, 2680 (1987).
15. Skupsky, S. *et al. J. appl. Phys.* **66**, 3456 (1989).

Rum tale of replication

Andrew W. Murray

MOST eukaryotic cells make sure that rounds of DNA replication (S phase) alternate with rounds of chromosome segregation (mitosis). The importance of this rule is highlighted by the exceptions to it. In the meiotic cell cycle, one round of replication is followed by two successive rounds of chromosome segregation, whereas polyploid and polytene cells arise from endoreduplication cycles, in which successive rounds of DNA replication occur without chromosome segregation (a in the figure). In an exciting advance, described on page 236 of this issue¹, Moreno and Nurse have identified the *rum1*⁺ gene of fission yeast as an important regulator of DNA replication: overexpression of *rum1*⁺ allows cells to endoreduplicate, while deletion of the gene allows them to undergo two successive mitoses without an intervening S phase.

Moreno and Nurse identified *rum1*⁺ by looking for fission-yeast genes that would perturb the cell cycle when they were overexpressed from a strong conditional promoter. Overexpression of *rum1*⁺ generated giant cells that contained many times the normal haploid content of DNA. Moreno and Nurse showed that this phenotype arose from multiple discrete rounds of DNA replication by investigating the interaction between *rum1*⁺

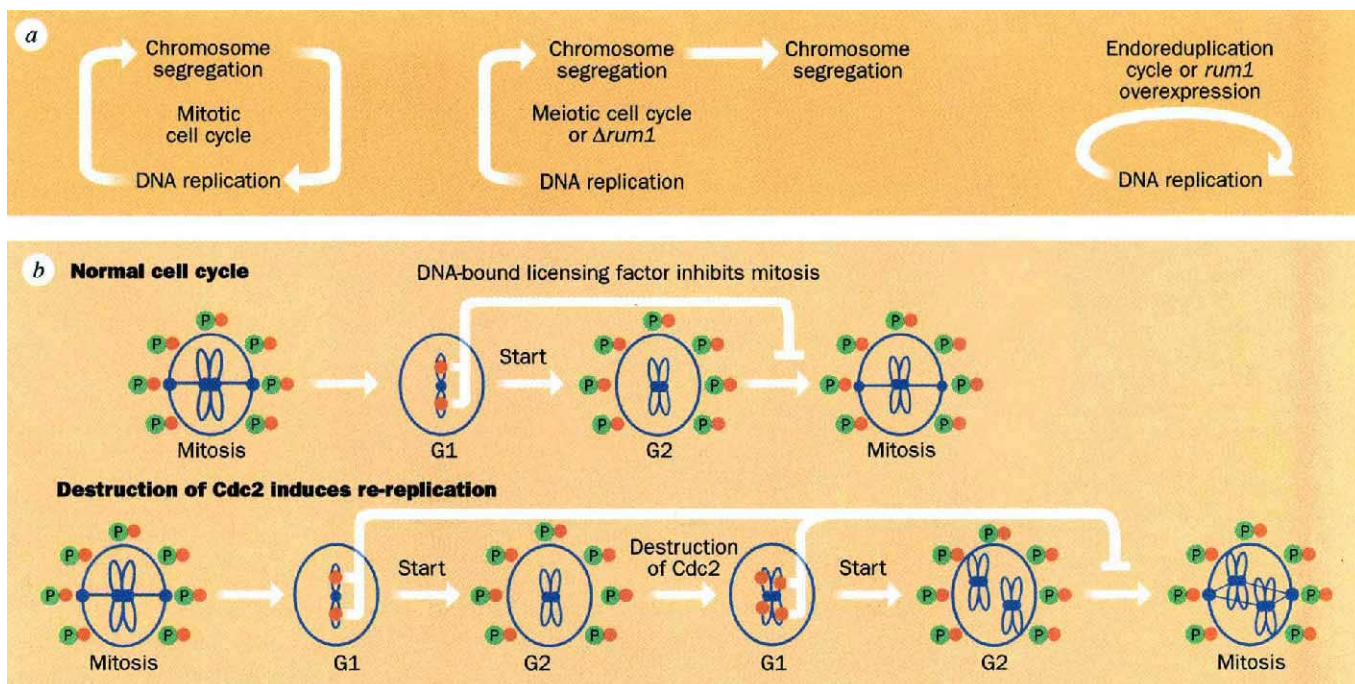
overexpression and mutations that prevent passage through mitosis or Start (the transition in the cell cycle which commits cells to entering S phase). Mutations that block entry into mitosis have no effect on *Rum1*-induced over-replication, but those that prevent cells from passing through Start prevent over-replication. This finding shows that *rum1*⁺ overexpression allows successive Starts without any intervening mitoses.

Deleting *rum1*⁺ abolishes the variable interval between the end of mitosis and Start, with the result that $\Delta rum1$ cells replicate their DNA as soon as they leave mitosis. If Start is blocked by the temperature-sensitive *cdc10*^{ts} mutation, the $\Delta rum1$ cells can still enter mitosis and divide, even though they have not gone through S phase. If the $\Delta rum1$ cells are arrested within S phase, however, they fail to enter mitosis. These results suggest that two different mechanisms control mitosis in response to the status of DNA replication. The first, revealed by deletion of *rum1*⁺, prevents cells that contain unreplicated DNA but have not passed Start from entering mitosis. The second, identified by a different set of mutants²⁻⁵, prevents cells that have passed Start, but have not completed S phase, from entering mitosis.

How does *Rum1* protein regulate the cell cycle? Moreno and Nurse base their explanation on earlier investigations into the role of Cdc2 — the master regulator of the cell cycle — in the control of DNA replication. These showed that conditions leading to the destruction of all of the Cdc2 in a G2 cell allowed that cell to undergo Start and replicate its DNA again within the same cell cycle⁶. These results prompted the hypothesis that Cdc2 exists in two different forms, an S form that allows cells to go through Start and replicate their DNA, and an M form that allows them to pass through mitosis and segregate their chromosomes.

In this scheme, *Rum1* is envisaged to play some critical part in the interconversion of the forms of Cdc2, such that overexpression of *Rum1* locks Cdc2 in the S form, and the absence of *Rum1* allows Cdc2 to remain in the M form after the end of mitosis. Perhaps the S and M forms are simply the association of Cdc2 with different cyclins? The S form would be Cdc2 bound to the G1 cyclins that help propel cells through Start, while the M form would be Cdc2 bound to the mitotic cyclins.

An alternative explanation of the alternation of Start and mitosis appeals to the notion of a licensing factor that is required for DNA replication. Experiments on the early embryonic cell-cycle of frogs led to the idea that replication depends on a licensing factor that binds to the DNA during mitosis and is then con-



a, Comparison of the normal cell cycle with cell cycles in which DNA replication and chromosome segregation do not alternate with each other. b, A model for regulation of mitosis by licensing factor. In yeast cells, licensing factor is phosphorylated and excluded from the nucleus during G2 and mitosis. The inactivation of Cdc2 at the end of mitosis leads to dephosphorylation of licensing factor and its entry

into the nucleus. The nuclear licensing factor can stimulate DNA replication and activates a pathway that prevents the induction of mitosis. Destruction of Cdc2 during G2 allows licensing factor to enter the nucleus. As a result, when Cdc2 activity is restored the cells cannot go through mitosis and instead pass Start for a second time and re-replicate their DNA.

sumed during S phase^{7,8}. In this model, the next round of replication in higher eukaryotes cannot occur until the nucleus has broken down, allowing the chromosomes to bind fresh licensing factor.

In budding yeast there is very good evidence for a licensing factor that controls, not DNA replication, but mating-type switching, an event whose cell-cycle regulation is similar to that of DNA replication. This licensing factor, the Swi5 protein, is made during G2 but is held in the cytoplasm because it is phosphorylated by Cdc28 (the budding yeast equivalent of Cdc2). At the end of mitosis, the inactivation of Cdc28 leads to the removal of the phosphate groups, allowing Swi5 to enter the nucleus, which like that of fission yeast has remained intact throughout mitosis⁹. Only when cells pass through Start is the Swi5 within the nucleus allowed to trigger the events that will lead to mating-type switching. Much weaker evidence supports the notion that a family of proteins, whose prototypic member is the budding yeast Cdc46 protein, act in a similar way to license DNA replication^{10,11}.

Could licensing factor in the nucleus regulate mitosis as well as DNA replication? In this model (*b* in the figure), the presence of licensing factor inside the nucleus prevents cells from entering mitosis. Like Swi5, licensing factor is made during G2, but phosphorylation by Cdc2 keeps it in the cytoplasm. The inactivation

of Cdc2 at the end of mitosis allows licensing factor to enter the nucleus and keeps mitosis from occurring before the beginning of DNA replication. Because licensing factor is consumed during S phase and can only re-enter the nucleus at the end of mitosis, this scheme allows cells to pass Start only once between successive mitoses. In addition, it provides a simple explanation for the re-replication of DNA after the removal of Cdc2 from a G2 cell⁶. The loss of Cdc2 allows licensing factor to enter the nucleus at an inappropriate point in the cell cycle (see figure). As a result, when Cdc2 is resynthesized the licensing factor in the nucleus prevents the onset of mitosis, allowing another Start and S phase to occur.

In this view of events, Rum1 may act as an inhibitor of Cdc2 that is activated by the presence of licensing factor in the nucleus. A recent News and Views article¹² discusses a number of protein inhibitors of Cdc2 and related cyclin dependent kinases (Cdk), some of which preferentially inhibit particular Cdk-cyclin complexes^{13,14}. In G1 the presence of licensing factor in the nucleus would activate Rum1, leading to a strong inhibition of Cdc2-mitotic-cyclin complexes that would prevent mitosis, and a weak inhibition of Cdc2-G1-cyclin complexes that would delay Start. Once licensing factor was consumed during DNA replication, the activity of Rum1 would decline, allowing the Cdc2-mitotic-cyclin com-

plexes to drive entry into mitosis and exclude licensing factor from the nucleus. The overexpression of Rum1 would inhibit the activity of this form of Cdc2 thus preventing mitosis and allowing licensing factor to enter the nucleus, triggering a second round of DNA replication.

It remains to be seen which of these models for the function of Rum1 is correct. But elucidating the biochemical function of this protein promises to yield further insights into the regulation of DNA replication. □

Andrew W. Murray is in the Department of Physiology, University of California at San Francisco, San Francisco, California 94143, USA.

1. Moreno, S. & Nurse, P. *Nature* **367**, 236–242 (1994).
2. Al-Khodairy, F. & Carr, A. M. *EMBO J.* **11**, 1343–1350 (1992).
3. Rowley, R., Subramani, S. & Young, P. G. *EMBO J.* **11**, 1335–1342 (1992).
4. Enoch, T., Carr, A. M. & Nurse, P. *Genes Dev.* **6**, 2035–2046 (1992).
5. Kelly, T. J. *et al.* *Cell* **74**, 371–382 (1993).
6. Broek, D., Bartlett, R., Crawford, K. & Nurse, P. *Nature* **349**, 388–393 (1991).
7. Blow, J. J. & Laskey, R. A. *Nature* **332**, 546–548 (1988).
8. Leno, G., Downes, C. S. & Laskey, R. A. *Cell* **69**, 151–158 (1992).
9. Moll, T., Tebb, G., Surana, U., Robitsch, H. & Nasmyth, K. *Cell* **66**, 743–758 (1991).
10. Hennessy, K. M., Clark, C. D. & Botstein, D. *Genes Dev.* **4**, 2252–2263 (1990).
11. Chen, Y., Hennessy, K. M., Botstein, D. & Tye, B. K. *Proc. natn. Acad. Sci. U.S.A.* **89**, 10459–10463 (1992).
12. Nasmyth, K. & Hunt, T. *Nature* **366**, 634–635 (1993).
13. Gu, Y., Turck, C. W. & Morgan, D. O. *Nature* **366**, 707–710 (1993).
14. Harper, J. W., Adam, G. R., Wei, N., Keyomarsi, K. & Elledge, S. J. *Cell* **75**, 805–816 (1993).

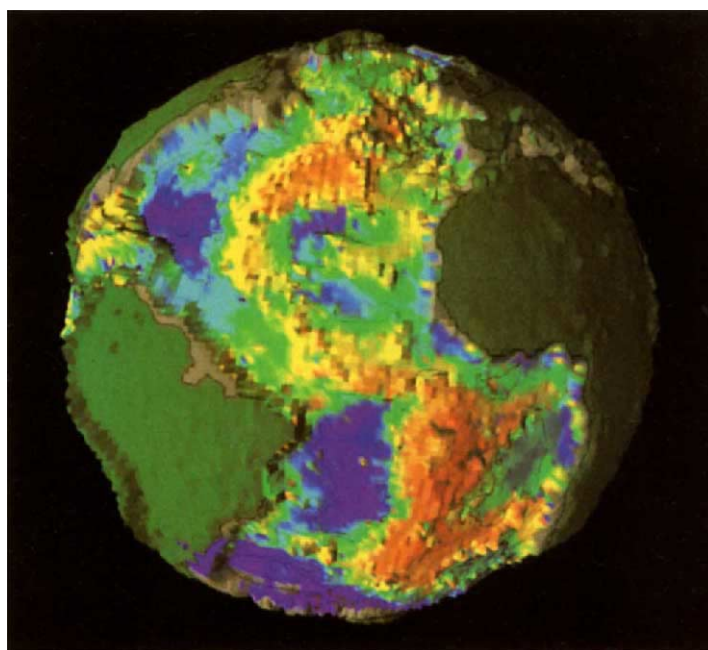
CARBON CYCLE

Compensation in the deep ocean

FOR calcium carbonate to dissolve in the oceans, pressure must be high, and the water must therefore be deep. In shallower regions, calcite is deposited on the sea bed, as shown in this computer simulation of calcite loading on and around the Mid-Atlantic Ridge. But the specific depth at which calcite first dissolves, and thus the area of the sea bed coated with calcite at any one time, depends on the total calcite concentration in a clever balancing act by the ocean. If the ocean receives more excess calcite via weathering, or some other source (such as changes in coral growth rates), carbonate

chemistry works to increase the dissolution depth, so that more calcite can be deposited on the sea floor and buried by sediments to maintain a steady state.

On page 260 Archer and Maier-Reimer



D. Archer

trolled by ocean carbonate chemistry, and use a model that combines ocean circulation, carbon cycling and various sedimentary processes to try to test what sort of changes in calcite deposition would be required to account for the glacial change.

One suggestion that they test is the effect of a change in the relative deposition rates of organic carbon and calcite to the deep sea — organic carbon degradation in the sediments changes the pore-water chemistry, which in turn encourages calcite to dissolve and sets in train the whole compensatory mechanism. They find that de-

creasing the rate of calcite deposition by just 40 per cent relative to organic carbon deposition would be enough to account for the entire glacial CO₂ decrease.

Gabrielle Walker

The variable ocean

Thomas F. Stocker

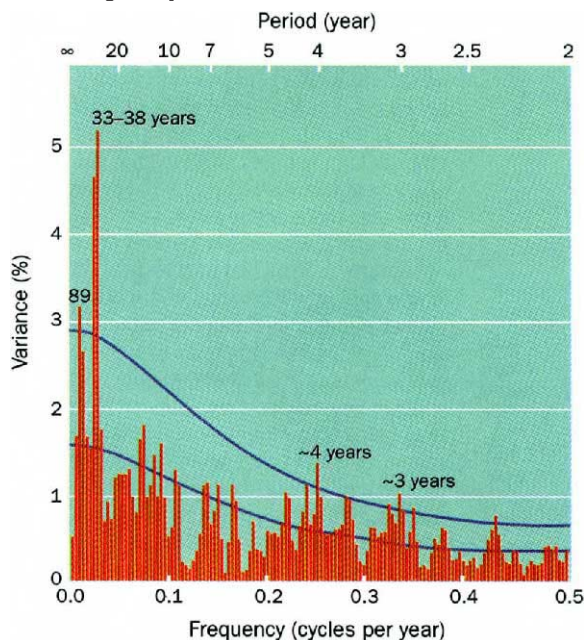
NATURAL variability in the climate system still hampers the unequivocal detection of increasing air temperatures as the emission of CO₂ into the atmosphere continues. Understanding such fluctuations is thus a great challenge to climate researchers, three of whom — Delworth, Manabe and Stouffer¹ — have now found natural variability in a 600-year integration of their general circulation model simulating the present climate. Irregular

worth and colleagues go further by analysing the natural variability of the modern-day, quasi-steady state and, in doing so, learning more about how the conspicuous interannual and interdecadal variability seen in long-term climate records is generated. The broad-band variability in the North Atlantic over a 600-year run compares favourably with that of observations from Weather Ship Station I (59° N, 19° W) and has the typical red-noise character. Superimposed on this random component is a broad peak at the interdecadal timescale, and the model results suggest a new form of interaction between ocean and atmosphere, occurring mainly in the Atlantic Ocean northeast of the maritime provinces of Canada.

The modelled interdecadal oscillation has its most noticeable effect on the North Atlantic 'conveyor belt', which carries vast quantities of surface water towards the pole where it cools and sinks to form deep water. There, its peak-to-peak amplitude is about $2 \times 10^6 \text{ m}^3 \text{ s}^{-1}$, more than ten per cent of the long-term mean. Such fluctuations in the rate of deep-water formation influence the strength of the surface flow into the sinking regions and with it the advection of heat and salt. By relating time series of sea surface temperature to a lagged time series of the conveyor strength, Delworth *et al.* demonstrate the succession of processes that are at

the heart of the simulated variability.

The schedule is as follows. When the thermohaline circulation is weak, decreased advection of lower-latitude warm and saline waters into the central region of the North Atlantic generates a pool of anomalously cold and fresh water. As the impact on density of the thermal anomaly is about three times stronger than the haline, the anomaly induces a geostrophically controlled cyclonic circulation at the surface, exchanging colder, fresher waters near the northeast coast of America with warmer, saline mid-ocean waters. This anomalous buoyancy-driven gyre is actually superimposed on the upper branch of the time-mean conveyor belt which flows northeast at that latitude and



Statistical methods allow the reconstruction of summer temperatures from tree-ring widths of pine trees. Spectral analysis of a 398-year series of this climate proxy from the northern regions of Fennoscandia shows there to be strong variability on a timescale of 30–40 years. The lower line represents red noise with equal variance and the upper is the 95 per cent confidence limit (after ref. 8; courtesy of K. R. Briffa). Natural fluctuations of the thermohaline circulation in the North Atlantic as found in a coupled atmosphere–ocean general circulation model¹ may well induce cycles of temperature in northwestern Europe.

but organized 40- to 60-year oscillations about a mean state are localized in the northwest Atlantic region and are associated with a subtle interplay between the thermohaline circulation of the ocean and the advection of heat and salt. The timescale of 40–60 years is especially interesting because it coincides with the typical timescales of anthropogenic climate modification and may hide changes for some years to come, unless advanced statistical methods are used².

The model, a three-dimensional simulation with atmosphere–ocean coupling developed at NOAA's Geophysical Fluid Dynamics Laboratory, is one of the standard tools in climate research to assess the consequences of global warming³. Del-

Time of a whale

THE name says it all: *Ambulocetus natans* could only be an all-walking, all-swimming intermediate between modern whales and their long-extinct terrestrial forebears. As J. G. M. Thewissen *et al.* show in the latest issue of *Science* (263, 210–212; 1994), the 50-million-year-old creature was about the size, weight and shape of a large sea lion. It had large hands and feet, but short, robust hind- and forelimbs, and the authors suggest it swam by combining movements employed by extant seals, otters and whales. Although this creature will be hailed as a 'missing link', systematists will argue about its status as a primitive whale rather than a derived mesonychid condylarth (extinct ungulates, taken to be most closely related to whales) — and, indeed, about the features that define whales as a group.

Far sight

DENSITY variations in the Earth's interior are largely derived after much wrestling with seismological data sets. At last month's meeting of the American Geophysical Society, C. Kuo (University of California, Berkeley) and collaborators proposed a very different way of tackling the question — they calculate that different densities should show up in measurements of neutrino absorption. High-energy neutrinos are needed, as the absorption probability rises with energy: neutrinos with energies of about 1 TeV have about an even chance of interacting with a nucleus as they pass through the Earth. By the end of this year, Kuo *et al.* hope enough results will have emerged from the detectors DUMAND II, in the deep ocean off Hawaii, and AMANDA, buried in ice in the Antarctic, to test their idea. This could be the nearest that geoscientists will come to holding the Earth up to a light and squinting at the shadows.

Stress relief

BOTANY meets biochemistry meets evolutionary biology meets genetic engineering in a report by A. D. Hanson *et al.* (*Proc. natn. Acad. Sci. U. S. A.* 91, 306–310; 1994). The authors looked at the occurrence in the Plumbaginaceae plant family of glycine betaine and four related compounds, which help protect organisms against the effects of drought or salinity. All species (thrift, for instance) of the Plumbaginaceae are resistant to osmotic stress, but they occur in a variety of environments. It seems that glycine betaine production is the basal condition, and from their analyses Hanson *et al.* surmise that each of the other compounds may well have further adaptive purposes; choline O-sulphate, for instance, may detoxify excess sulphate as well as offer osmoprotection. The options for crop improvement with recombinant DNA technology grow ever more subtle.

enhances it. As the associated increased salt transport brings denser water into the sinking region and so helps the conveyor intensify, more heat is advected from lower latitudes by the same mechanism. This erodes the cold anomaly and eventually creates a warm pool in the same region. Geostrophy now requires an anticyclonic gyre which weakens the thermohaline circulation by carrying fresher and colder waters in the mid-ocean southwards, against the direction of the time-mean conveyor belt.

Budget calculations in this region indicate that the fluctuation is mainly due to periodically varying meridional fluxes of heat and salinity. Ocean-to-atmosphere exchange is important for heat (it slowly removes temperature anomalies) but not for salt; surface fluxes of fresh water vary only slightly during a cycle. This indicates that the fluctuation is mainly oceanic and could, with suitable surface boundary conditions, be simulated in ocean-only models. But classical 'mixed boundary conditions' (where heat fluxes are proportional to temperature anomalies, freshwater fluxes are constant) will not do the trick, as it is known that they thermally couple the atmosphere too tightly to the ocean. As Delworth and colleagues show, the persistence of sea-surface-temperature anomalies is a central part of the proposed mechanism and also acts to stabilize the North Atlantic circulation.

Although the fluctuations in the model are initiated at the sea surface, their effects can be traced both in the deep Atlantic and in the atmosphere of the Northern Hemisphere. Salinity and temperature profiles show changes that reach beyond 3 km depth in the sinking region between 52° and 72° N in the Atlantic. Surface air temperatures also follow the oceanic oscillation, and distinct anomaly patterns are expected, especially in winter months when a stronger thermohaline circulation transports enough heat northward to reduce ice cover. During that phase of the oscillation a negative pressure anomaly develops, centred southeast of Greenland, bringing a warming of about 1 °C over the northern North Atlantic and a significant cooling of about 1.5 °C over the Labrador Sea.

The modelled oscillation is distinctly irregular, a common feature in nonlinear dynamical systems. Although the first 200 years of the integration show a quasi-periodic cycle of 40–50 years, the periods are longer in the following 400 years. It seems that some preconditioning of the ocean at a particular location is required, and that the fluctuation itself may remove or add to it. Other model calculations also point at a centre of action in the northern North Atlantic and Labrador Sea^{4,5}, and they also show oscillations on timescales of 10–40 years, although these are due to somewhat different mechanisms.

What adds weight to the model results is that observational studies indicate distinct and coherent changes in the ocean and atmosphere in this century. A hard look at global-mean temperature records reveals interdecadal (60–70 year) cycles and increased variability in the region of the North Atlantic². A long-term cooling trend from the late 1950s to early 1970s, in the top 1,500 m of the Atlantic Ocean at 34° N, follows patterns of sea surface temperature and salinity anomalies^{6,7} that are similar to the modelled differences between times of strong and weak thermohaline circulation. Waters in the central Labrador Sea, too, have become cooler and fresher at intermediate depths during the past 25 years⁸, but with the coarse resolution of present models a direct comparison is not yet possible.

Perhaps the most intriguing piece of evidence for interdecadal cycles comes from a 1,480-year time series of summer temperatures reconstructed from tree-ring widths of pine from northern Fennoscandia⁹. The past 400 years of this proxy indicator show significant variance in the band of 30–40 years (see figure). Spectral analysis of sub-intervals indicates that these periods themselves are variable¹⁰, as in the numerical simulation. Delworth and colleagues' model could provide a natural mechanism for such cycles (previous interpretations have had to invoke some nonlinear interaction between solar forcing and atmospheric temperatures).

Given the paramount importance of the thermohaline circulation and its variability for the climate system, what should the next steps be? Enlarging the observational data set in key regions such as the North Atlantic and Labrador Sea area must be one priority. And numerical models must be refined to give a better representation of how deep water masses are formed and distributed, plus improved compatibility of atmospheric and oceanic heat, freshwater fluxes, and air–sea interaction. We will then gain more of a feel for the robustness of such cycles. □

Thomas F. Stocker is in the Physics Institute, University of Berne, Sidlerstrasse 5, 3012 Berne, Switzerland.

1. Delworth, T., Manabe, S. & Stouffer, R. J. *J. Clim.* **6**, 1993–2011 (1993).
2. Schlesinger, M. E. & Ramankutty, N. *Nature* (in the press).
3. *The IPCC Scientific Assessment*, 366 (WMO-UNEP, Cambridge Univ. Press, 1990).
4. Weisse, R. *et al.* *J. geophys. Res.* (in the press).
5. Weaver, A. J., Aura, S. M. & Myers, P. G. *J. geophys. Res.* (in the press).
6. Kushnir, Y. *J. Clim.* **7**, 141–157 (1994).
7. Levitus, S. *J. geophys. Res.* **94**, 6091–6131; 9679–9685 (1989).
8. Lazier, J. R. N. in *Natural Climate Variability on Decade-to-Century Time Scales* (National Research Council, in the press).
9. Briffa, K. R. & Schweingruber, F. H. in *Climate Since A.D. 1500* (eds Bradley, R. S. & Jones, P. D.) 366–392 (Routledge, London, 1992).
10. Briffa, K. R. *et al.* *Clim. Dynam.* **7**, 111–119 (1992).

Electric squeezer

DAEDALUS has been musing on the pinch effect, which applies a sort of 'electric pressure' to the surface of a current-carrying wire. He argues that molecules adsorbed on the surface should feel this pressure too. Many catalytic chemical processes depend on high pressures. The chemical industry is overburdened with massive reaction vessels, heavy-gauge pipework, compressors and so on. Daedalus wants to use the pinch effect to dodge all this expensive complexity.

Imagine, he says, a wire suspended in a mixture of hydrogen and nitrogen, and coated with ammonia-synthesis catalyst. Pass a current through the wire, and its catalytic surface and the gas molecules adsorbed on it would feel the pinch pressure. A sufficiently heavy current would give a surface pressure high enough to combine the adsorbed gases into ammonia and to liquefy the product. If the wire were porous with a central duct, the liquid ammonia would be forced inwards as fast as it formed, and would run out of the ends of the duct in a steady stream. All the desperate high-pressure engineering of conventional synthesis would be neatly evaded.

The same trick would work in other pressure technologies, even outside the chemical industry. A porous pinch-effect wire in a refrigerant vapour could act as a silent freezer–compressor. In air it would form a simple, elegant and continuous dehumidifier, and an extractor of condensable polluting vapours. In both cases the wire would need a heat sink to the outside to remove both ohmic heating and latent heat of condensation: a small complication to an otherwise elegantly simple design.

Daedalus also reckons that the principle could revolutionize electric lighting. The old carbon-filament lamp rapidly deteriorated as its filament evaporated. The obvious answer is to drive so heavy a current through the filament that its surface pinch pressure exceeds the vapour pressure of the carbon. Evaporation would then be frustrated. Daedalus calculates that such a filament could be taken right up to 5,000 K, the melting point of carbon and almost as hot as the surface of the Sun. This brilliant new pinch-pressurized lamp will work on d.c. only, because the filament would evaporate during the zero current intervals in a.c. For the same reason, it will be restricted to tunnels, factories and security installations, where lamps run all the time: once lit, a pinch-pressurized lamp cannot be turned off. The instant the current ceases and the pinch pressure relaxes, the superheated filament will explode.

David Jones

Destruction of toxic materials

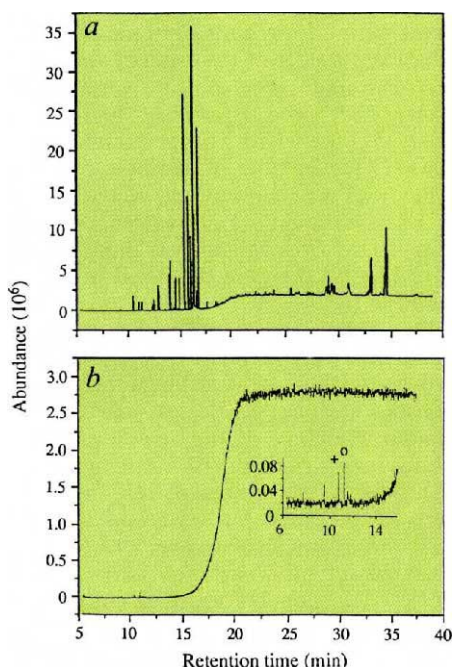
SIR — Increased awareness and understanding of the hazards to health and the environment of many synthetically produced chemicals have led to the need for development of a safe means for their effective disposal. Because of their high toxicity and persistence, halogenated organic compounds and other organochlorine pesticides such as DDT, PCBs, aryl halides and dioxins are of particular concern. Current methods of disposal¹ involving high-temperature chemical reactions can be accompanied by unwanted toxic emissions or by-products, which may involve unacceptable risks to the environment. We have developed a low-temperature process² involving mechanochemical reactions³ induced by mechanically milling the toxic material with a suitable reactant in a closed vessel. Such reactants may include reductants which break down the entire molecule or react selectively to remove chlorine. Laboratory measurements of the mechanochemical reaction of DDT, PCB and chlorobenzene with magnesium, calcium and calcium oxide show the destruction of the organochlorines without the formation of toxic by-products.

We milled individual samples of chlorinated organic compounds, including DDT, PCB (Aroclor 1254) and chlorobenzene with Ca, Mg, Fe or CaO as a reactant in a laboratory vibratory mill (SPEX 8000). Each test consisted of milling a total of 5–10 g organochlorine and the reactant using 12-mm diameter steel balls in a sealed steel vial. The figure shows a total ion chromatogram (TIC) from gas chromatography-mass spectroscopy (GC-MS) of compounds present after milling a sample of DDT and Mg for 3 and 6.5 h. Chlorinated break-up products and products of partial dechlorination of DDT were present after 3 h milling (*a*) in the figure, retention time *t* = 10–30 min). Partially dechlorinated oligomer-type materials of DDT (*t* = 31–35 min) were also present. The presence of chlorinated compounds was not evident from analysis of the TIC from GC-MS of the sample milled for 6.5 h (*b*, *t* = 10–12 min). We found very small quantities (low p.p.m.) of dechlorinated products, such as diphenylmethane and diphenylethanes (inset in *b*). These compounds were not present in the DDT used for milling, so must have formed during milling.

The levels of DDT and related compounds DDD (DDT minus Cl) and DDE (DDT minus HCl) and the levels of chlorobenzene and PCB determined by dual-column gas chromatography-electron capture detection and expressed as the fraction in p.p.m. of the original organochlorine compounds remaining after being milled for 12 hours, are summa-

rized in the table. For each reactant, the results for the three organochlorines were similar even though both the PCB and chlorobenzene were in liquid form whereas the DDT was solid. Mg, Ca and CaO were all highly effective in dechlorinating the organochlorine reactants. When Fe was used, milling for 12 h resulted in incomplete dechlorination.

The importance of mechanochemical activation is that, with suitable choice of reactants, the reaction can occur at low temperatures in a sealed environment to form safe reaction products requiring only minimal additional processing. Further, the process is based on well-established ball-milling technology and can be scaled



GC-MS chromatograms of samples of DDT milled with Mg: *a*, after 3 h milling; *b*, after 6.5 h milling. Cross, diphenylmethane; circle, diphenylethane. Measurements were carried out after extraction into *n*-pentane using a HP 5890 II GC-MSD 5790 instrument fitted with a DB-5 column of 30 m length, inside diameter 0.25 mm and film thickness 0.25 μ m. The carrier gas was helium at 70 kPa. The temperature programme was 40 °C for 2 min, then 15 °C per min to 310 °C and held for 20 min.

MILLING EXPERIMENTS WITH VARIOUS REACTANTS			
Reactant	Remaining organochlorine (p.p.m.)		
	DDT	Chlorobenzene	PCB
Mg	4	<1	260
Ca	<1	50	1,400
Fe	36,000	280	120,000
CaO	4	7	5

Values under 'DDT' also refer to DDD and DDE.

to any size. Mobile mills mounted on road or rail vehicles may be used for on-site processing, thus eliminating the environmental hazards associated with the transport of toxic materials. CaO would be the preferred reactant because it is cheap and the reaction products can be disposed of safely.

S. A. Rowlands

A. K. Hall

P. G. McCormick

Department of Mechanical and Materials Engineering,

R. Street

*Department of Physics,
University of Western Australia,
Nedlands, WA 6009, Australia*

R. J. Hart

G. F. Ebell

*Environmental Chemist Laboratory,
Chemistry Centre of WA,
East Perth, WA 6004,
Australia*

P. Donecker

*CRA Advanced Technical Development,
Technology Park,
Bentley, WA 6102, Australia*

1. Murena, F. Famiglietti, V. & Gioia, F. *En. Prog.* **12**, 231 (1993).

2. Australian Patent Application PL6474 (1992).

3. Heinicke, G. *Tribochemistry* (Hanser, Munich, 1984).

High jump

SIR — Jha's note (*Nature* **365**, 398; 1993) about timing devices in athletic events reminds me that two different philosophies are used in judging competitions. In the first category are running, throwing and long-jumping events, in which times or distances achieved are measured as exactly as possible by existing technology.

The second category comprises high-jumping and pole-vaulting. Here the bar is either cleared or not, and its final height is measured exactly for the record. Although the winning jumper usually clears the bar with distance to spare, he/she receives no credit for that extra distance, other than anecdotally.

We could determine the actual height cleared for the record. We could raise a curtain of laser light along and above the bar shining from one verticle support to the other, and we could measure the height cleared by recording the brief attenuation of the light as the athlete's body or clothing passes through the beam. The bar would remain as a psychological aid to the jumper, and as the traditional means of determining the winner.

But high-jumping is an especially pure, simple and human contest, free of devices aiding the jump or determining the result. Perhaps we should just leave it that way.

James Johnson

*Curams Kennels,
Westport,
Canada KOG 1X0*

A view from the Sunda arc

SIR — Plank and Langmuir¹ argue that the sedimentary flux down subduction zones of elements like Ba, Sr, K, Rb, Cs, La, Th and U is reflected by the abundances of these elements in the volcanic outputs from the arc once the effects of melting are taken into account. They illustrate their argument with data for eight volcanic arcs, including the Sunda arc, for each of which they compare an average basalt composition with a calculated sediment flux based on the rate of subduction and on the bulk composition of a representative sediment section.

For the Sunda arc, Plank and Langmuir calculate an average basalt composition with about 1% K from data for six central Javanese volcanoes. But it seems pointless to calculate an average Sunda basalt when it is known that the abundances of K-group elements in mafic volcanics in Sunda volcanoes vary along the arc² as well as across it³; they are low in west Javanese volcanoes but increase going eastwards as far as central Java, then at about 112°E revert to relatively low contents but again increase eastwards as far as Sumbawa. Therefore, the abundances of K (which in the Sunda arc are well-correlated with Rb, Sr, Nb, Th and La abundances) in mafic (6% MgO) magma compositions legitimately selected as representative of various sectors of the Sunda arc could vary, using Plank and Langmuir's approach, from 0.5 to 3.00% K depending on which sector of the arc one selects the volcanoes. How can Plank and Langmuir explain this along-arc variation as a consequence mainly of sediment flux?

I also wonder how good the correlation really is between K/Na and Ba/Na values which Plank and Langmuir calculate from their average basalt compositions, and their calculated sediment fluxes of K and Ba. This correlation is the main evidence they cite in support of their general argument (Fig. 3 of ref. 1). For the Sunda basaltic data point, Plank and Langmuir averaged D. J. Whitford's data for six central Javanese volcanoes to get values of 0.34 for their K/Na parameter and 90 for their Ba/Na parameter. But, for their six selected Javanese volcanoes, the values of the Ba/Na parameter include two at 35 and 36, and two at 141 and 206, so that four of the six volcanoes fall outside the range of their averages for all eight arcs. Similarly, the values of their K/Na parameter include two at 0.13 and 0.16, and two at 0.50 and 0.66, again outside the range of their averages for the eight arcs.

Not only are Plank and Langmuir placing a lot of faith in their selection of the volcanoes as representative and in their averaging approach, but the range of

values they averaged is actually also representative of real variation between different parts of the arc, so that their own approach suggests that in the Sunda arc there might instead be a very poor correlation between basalt composition and sediment flux, at least for Ba and K.

It seems to me that Plank and Langmuir's explanation fails to account for one of the most interesting geochemical features of the world's most active arc. I would be interested to know how it looks when viewed from the seven other arcs.

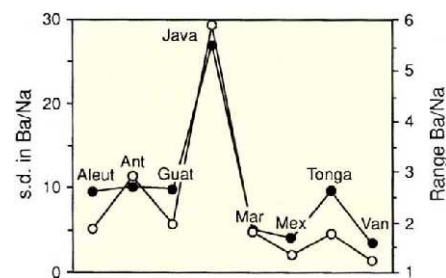
R. Varne

*Département de Petrologie,
Université Pierre et Marie Curie,
75252 Paris, Cedex 05, France*

Present address: Department of Geology, University of Tasmania, Hobart, Australia 7005.

PLANK AND LANGMUIR REPLY — We thank Varne for focusing attention on an important aspect of the Sunda volcanics: the extreme geochemical variability. There are three significant issues he raises: (1) the validity of the Sunda average; (2) the validity of averages for the other arcs we considered; and (3) the origin of variability among volcanoes from a single arc. We shall address these issues here briefly; they are discussed in more detail in ref. 4.

Our general goal¹ was to test whether sedimentary input affects volcanic output. We would like to investigate this relationship at all length scales, but current limited sampling of the sea floor near trenches restricts us to examining relationships at the global scale. Thus we focused on inter-arc variations, which required us to calculate averages for the volcanic output at each arc. Large variations within populations does not invalidate comparison of the means of the populations. We share Varne's concern about the Java average; as shown by the ± 1 sd error bars in Fig. 3 of ref. 1, it is the most uncertain of the averages and deserves scrutiny. For the Sunda arc,



Uncertainty in the mean Ba/Na of eight volcanic arcs, expressed as 1 s.d. of the mean (solid symbols) and as the total range (high value/low value) (open symbols). Java is clearly anomalous. Other arcs include the Eastern Aleutians (Aleut), Lesser Antilles (Ant.), Guatemala/El Salvador, (Guat), Marianas (Mar), Mexico (Mex), Tonga and Vanuatu (Van).

tectonic constraints and data availability limited us to volcanism on Java, where continent collision has not yet begun, and to six volcanoes for which geochemical data exist on mafic samples⁵. The variation in these six Javanese volcanoes is as large as the global range of all arc volcanoes, so calculating an average for Java depends strongly on how representative these volcanoes are of the total volcanism on Java (which include more than 35 active volcanoes⁶). To calculate the most accurate average possible, we classified all the active volcanoes on Java as low-K, medium-K or high-K (on the basis of available data), and calculated a weighted mean of these classes using mafic data from the six volcanoes (two of which fell into each class). We agree that the average is sensitive to the weighting, but believe that the average is a fair representation of the average volcanic flux on Java, and is the appropriate value with which to compare to the average sediment input. The Java average is not significantly affected (less than 6% relative) by including nearby volcanoes on Bali, and so is not strongly dependent on the length of the arc sector chosen.

For most arcs, the individual volcanoes define a unique region in geochemical space (Fig. 2 of ref. 1), and a simple average of the volcanic data was sufficient to characterize the arc from a global perspective. Java is the exception, as is apparent from the attached figure. The correlations we presented in Fig. 3 of ref. 1 are not dependent on the Java point; the regression based on the other arcs is as good or better than the one that includes Java. Therefore Varne's concern about the validity of the averages from the other seven arcs should be assuaged.

We suggest, however, that even the Java average reflects its sediment input. Javanese volcanics possess robust geochemical features that are shared by the Indian Ocean sediments: both are global end-members in Th/Sr, and the Ba/Th ratio for Java volcanics is distinctively low and fairly uniform. Other studies focused specifically on Indonesia have also found an important role for sediment recycling⁷⁻¹⁰.

1. Plank, T. & Langmuir, C. H. *Nature* **362**, 739–743 (1993).
2. Wheller, G. E., Varne, R., Foden, J. D. & Abbott, M. J. *J. Volcanol. Geotherm. Res.* **32**, 137–160 (1987).
3. Whitford, D. J., Nicholls, I. A. & Taylor, S. R. *Contrib. Mineral. Petrol.* **70**, 341–356 (1979).
4. Plank, T. thesis (Columbia Univ., 1993).
5. Whitford, D. thesis (Australian National Univ. 1975).
6. Simkin, T. *et al.* *Volcanoes of the World* (Smithsonian Institution, Washington, DC, 1981).
7. Whitford, D. J. & Jezek, P. A. *Geol. Soc. Am. Bull.* **93**, 504–513 (1982).
8. Stolz, A. J. *et al.* *Contrib. Mineral. Petrol.* **105**, 585–601 (1990).
9. Edwards, C. M. H. *et al.* *Nature* **362**, 530–533 (1993).
10. Vroon, P. Z. *et al.* *J. geophys. Res.* (in the press).
11. Klein, E. M. & Langmuir, C. H. *J. geophys. Res.* **94**, 4241–4254 (1989).
12. Langmuir, C. H. in *Caltech Plume Symp.* (Pasadena, 1991).

The final issue is the origin of local variability. It may have little to do with sediment fluxes. Volcanic data from other tectonic settings has shown us that different processes may be responsible for geochemical variations at different length scales. For example, local geochemical trends on the Mid-Atlantic ridge and some ocean islands are actually orthogonal to global trends, and appear to derive from different magmatic phenomena^{11,12}. Similarly, we can imagine many processes that could effect local, intra-arc variability: mantle heterogeneity, sediment heterogeneity, variable sediment or fluid delivery, temporal variations, etc. The origins of this local variability are an important and on-going topic of research. We hope that with expanded data sets, the patterns and origins of local variability in subduction zones will be revealed.

Terry Plank

Department of Geological Sciences,
Cornell University,
Ithaca, New York 14853, USA

Charles H. Langmuir

Lamont-Doherty Earth Observatory,
Columbia University,
Palisades, New York 10964, USA

A bioluminescent chaetognath

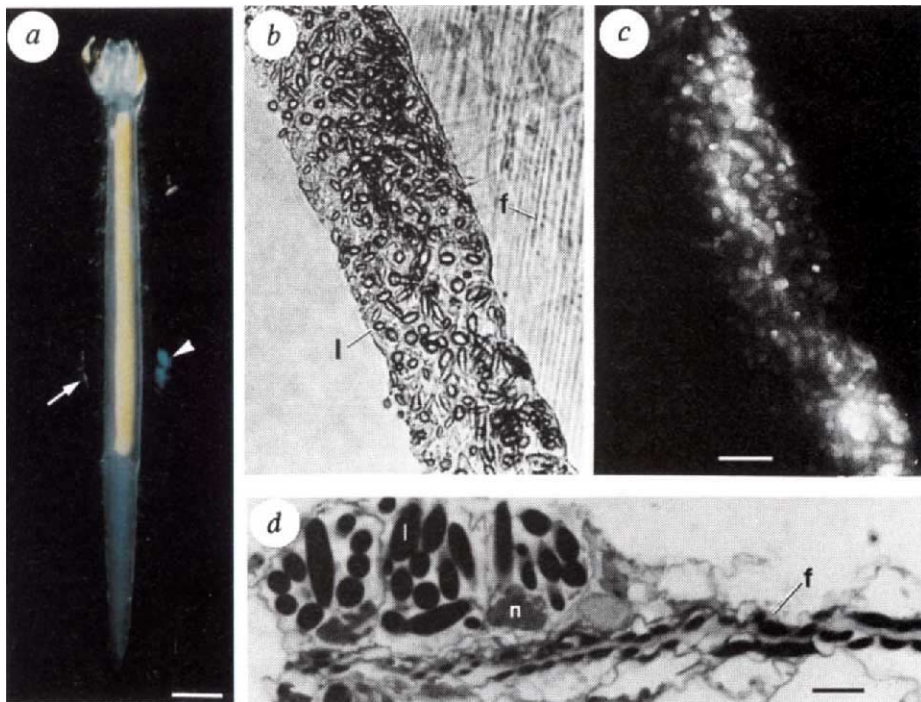
SIR — The ability to produce light is especially widespread among marine zooplankton, where the only major phylum without a luminous species has been thought to be the Chaetognatha¹.

These carnivorous arrow worms are found throughout the world's oceans, where their abundance approaches that of copepods². We have discovered bioluminescence in a well-known bathypelagic chaetognath. This is the first addition to the roster of luminescent phyla in more than 50 years³.

Chaetognath bioluminescence was first noted from the *Johnson-Sea-Link* submersible on cruises with E. A. Widder near the Bahamas Islands. On three dives, at 870, 850 and 760 m, a chaetognath with an orange-pigmented gut darted away, leaving a plume of luminescence.

We subsequently examined chaetognaths collected from the North Pacific Ocean on cruises with J. J. Childress and E. V. Thuesen. Luminescence was repeatedly evoked in specimens of *Caecosagitta macrocephala* (Fowler, 1905), a cosmopolitan species generally found at depths greater than 700 m (ref. 4) (a in the figure). Luminescence was not produced by *Eukrohnia fowleri*, the other common species with an orange gut, nor by nine other chaetognath species examined (identified by E. V. Thuesen).

In marine invertebrates luminescence is produced via bacterial symbionts, or



Sources of luminescence in the chaetognath *C. macrocephala*. a, Composite image showing the yellow luminous bodies (arrow) on the edge of the transparent anterior fins (not visible) and their blue emitted light (arrowhead). b, Light and c, fluorescence microscopy of the region indicated by the arrow in a. d, Cross section through the edge of the fin showing three large cells, each with several luminous organelles, attached above the supportive fin rays. l, Light-emitting organelle; n, nucleus; f, fin ray. Scale bars: a, 2 mm; b, c, 40 μ m; d, 10 μ m.

through the oxidation of a light-producing compound (luciferin), mediated by a calcium-activated photoprotein or by an enzyme (luciferase). Methanolic extracts of *C. macrocephala* contained coelenterazine, the luciferin of cnidarians, ctenophores, radiolarians, and some squid, fish, and crustaceans⁵. Aqueous extracts added to coelenterazine showed high levels of luciferase activity. (Substrates for the assays were provided by O. Shimomura and S. Inoue.) No calcium-activated photoprotein was detected, and an assay for bacterial luciferase performed by M. G. Haygood and G. Mowlds was also negative. These results indicate that light is produced by a coelenterazine-luciferase reaction, making Chaetognatha the seventh phylum known to use coelenterazine for its luminescence.

Light is produced at the midpoint of the body by large vacuolar cells on the edges of the anterior fins (a in the figure). Oval or fusiform membrane-bound inclusions within these cells (b, d) are autofluorescent to varying degrees (c), a trait that has been correlated with the location of luminescent sources in many taxa^{6,7}. The fluorescent subcellular bodies are likely sites for storage of the components of the luminescent reaction. Electron microscopy (not shown) indicates that most contain a dense paracrystalline matrix with small spherical inclusions. Others, even within the same cell, lack internal organization, possibly indicating organelles in which the luminescent

compounds have reacted.

Although the luminous bodies are visible to the naked eye and may even be visible in a published micrograph⁸, they have not been included in previous descriptions of this species⁸⁻¹¹. These omissions, as well as the failure to note bioluminescence, may be due in part to the poor condition of specimens examined in the past; the insulated closing cod end of the 10 m² Tucker trawl¹² used in this study recovers animals in excellent condition.

Chaetognaths are commonly allied with pseudocoelomates (= Aschelminthes) and with deuterostomes, but recent molecular work supports the view that they diverged near the roots of the Metazoa¹³. Based on this early divergence and on the uniqueness of the light-producing cells, it appears

1. Herring, P. J. *J. Biol. Chem.* **1**, 147–163 (1987).
2. Reeve, M. R. in *Marine Food Chains* (ed. Steele, J. H.) 168–189 (University of California Press, Berkeley, 1970).
3. Kanda, S. *Biol. Bull.* **77**, 166–173 (1939).
4. Pierrot-Bults, A. C. & Nair, V. R. in *The Biology of Chaetognaths* (eds Bone, Q., Kapp, H. & Pierrot-Bults, A. C.) 86–116 (Oxford University Press, 1991).
5. Campbell, A. K. & Herring, P. J. *Mar. Biol.* **104**, 219–225 (1990).
6. Herring, P. J. *et al. Mar. Ecol. Prog. Ser.* **94**, 297–309 (1993).
7. Morin, J. G. & Reynolds, G. T. *Biol. Bull.* **139**, 430–431 (1970).
8. Casanova, J. -P. *J. nat. Hist.* **26**, 663–674 (1992).
9. McLelland, J. A. *Gulf Research Reports* **8**, 145–172 (1989).
10. Alvarino, A. *Naga Report* **4**, 1–197 (1967).
11. Fowler, G. H. *Trans. Linn. Soc. London ser. 2. Zool.* **10**, 55–87 (1905).
12. Childress, J. J. *et al. Deep Sea Res.* **25**, 419–422 (1978).
13. Telford, M. J. & Holland, P. W. H. *Molec. Biol. Evol.* **10**, 660–676 (1993).

that luminescence evolved independently in *C. macrocephala*, although in some organisms luciferin can be obtained from food.

Because luminescence is normally produced in conjunction with an escape response, the cloud of light appears to function as a diversionary display. This adaptation highlights the importance of bioluminescence in the interactions among marine organisms and demonstrates the value of *in situ* observations for understanding life in the sea.

Steven H. D. Haddock

James F. Case

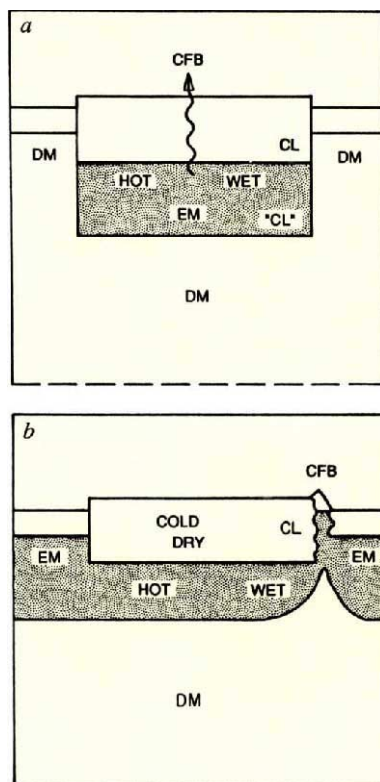
Marine Science Institute,
University of California,
Santa Barbara, California 93106, USA

Lithosphere and flood basalts

SIR — So-called hotspot magmas, including ocean island basalts and continental flood basalts, differ isotopically from mid-ocean-ridge basalts (MORB) and require a reservoir, or a component, that has been isolated from the MORB source for more than 2 billion years. It has been widely assumed that a suitable isolated reservoir might be the continental lithosphere.

Most authors refer to 'lithosphere' as the strong outer shell of the Earth. Geophysics¹ and rock physics² show that the silicates of the mantle have little strength at temperatures greater than 650 ± 100 °C or about half the absolute melting temperature. Some authors, however, refer to the thermal boundary layer as 'lithosphere' even though it extends to about 1,300 °C, near or above the silicate solidus³. The thermal boundary layer has a close to critical Rayleigh number and cannot remain attached to the plate or accumulate long-term isotope anomalies³ unless it is also isolated from the 'convecting mantle' by buoyancy^{4,5}. Only the shallow colder part of the thermal boundary layer can be considered rigid, or can maintain a long-term attachment to the plate or to an overlying continent. The use of the term 'lithosphere' for the thermal boundary layer is unfortunate, as it has led geochemists to believe that it is strong and can remain attached to ancient crust and its associated mantle (the rheological or real lithosphere) for long periods.

Both the lithosphere and the thermal boundary layer are colder than the rest of the mantle and are relatively thin. This makes it hard to see how large amounts of melt can be produced in a short period of time, as is characteristic of continental flood basalt provinces⁶. Gallagher and Hawkesworth⁷ therefore propose a "wet continental lithosphere" reservoir with a melting temperature ~ 500 °C lower than



a, The continental lithosphere (CL) hypothesis attributes continental flood basalts (CFB) to the lower part of the thermal boundary layer under continents, called by some the "lithosphere". This material has asthenosphere-like seismic velocity and viscosity and cannot be isolated from depleted mantle (DM) by its strength. It will flow laterally. b, The perisphere model attributes enriched components of hotshot magmas to a weak enriched mantle (EM) tapped at continental rifts or in the initial stages of DM upwelling. The strong lithosphere or plate is limited to regions colder than 650 ± 100 °C. Basalts from DM are not evident in the initial stages of rifting but become dominant at mature or rapidly spreading ridges.

that of dry silicates. The lower part of this proposed 'lithospheric' source has temperatures ranging from 1,000 to 1,500 °C. Such mantle has low viscosity and low seismic velocities. The presence of even trace amounts of water results in significantly lower creep strength than under strictly dry conditions. The strain rate of wet olivine² at 1,200 °C for a typical lithospheric deviatoric stress of 500 MPa is 10^{-8} s⁻¹. Lithosphere straining uniformly at this rate will experience a strain of unity in 3 years. Such high stresses cannot be maintained in hot mantle, which is thus better described as 'asthenosphere' (weak layer). The lower part of the thermal boundary layer cannot support even its own weight for long periods of time. It will also delaminate or deform irreversibly when subjected to compression or extension. It does not contribute to the strength of the lithosphere and deforms independently of it.

Observed seismic velocities in cratonic

lithosphere require a refractory olivine-rich composition⁸. Seismic shear velocities at 1,200 °C and 150 km depth in olivine-rich mantle are 4.6 km s⁻¹, and they will be even lower in wet mantle. Observed shear-wave velocities at this depth under ancient cratons are 4.78 km s⁻¹, which imply temperatures of ~ 600 °C (ref. 8). Non-cratonic mantle has lower velocities.

Thus, the physical properties calculated for wet lithosphere are asthenosphere-like. The conditions envisaged for continental 'lithosphere' will not be long maintained and the material will flow readily. Metasomatized or hydrous mantle is buoyant, particularly if it is basalt-depleted harzburgite, and therefore is likely to spread across the top of sub-lithospheric mantle (see figure). Fossil plume heads, another proposed source for flood basalts⁹, are even hotter than continental lithosphere and will suffer the same fate. In either case an enriched sub-lithospheric layer will result. Such a layer, which I have called the 'perisphere'¹⁰, will isolate the deeper MORB reservoir from recycling and fluids resulting from slab dehydration. In addition to these components it may also be rich in residual, small-melt fraction fluids^{4,11,12}. A global shallow enriched or metasomatized layer is intrinsic in some mantle evolutionary models^{4,8,11}. The enriched perisphere can eventually be pushed aside or depleted, just as has been proposed for continental lithosphere and 'fossil plume heads', to allow egress of depleted mantle or MORB melts therefrom. Attributing the source of flood basalts to continental lithosphere rather than the hot and weak sub-lithospheric mantle results from a semantic confusion between 'lithosphere' and 'thermal boundary layer'. A recent reappraisal of the geochemistry of flood basalts and continental lithosphere also favours a sub-lithospheric origin¹³.

Don L. Anderson

Seismological Laboratory,
California Institute of Technology,
Pasadena, California 91125, USA

- McNutt, M. K. *J. geophys. Res.* **89**, 11,180–11,194 (1984).
- Mackwell, S. J., Bai, Q. & Kohlstedt, D. L. *Geophys. Res. Lett.* **17**, 9–12 (1990).
- White, R. S. *J. Petrol. Spec. Vol.* 1–10 (1988).
- Anderson, D. L. *Earth planet. Sci. Lett.* **57**, 1–24 (1982).
- Anderson, D. L., Tanimoto, T. & Zhang, Y.-S. *Science* **256**, 1645–1651 (1992).
- Arndt, N. & Christensen, N. J. *geophys. Res.* **97**, 10,967–10,981 (1992).
- Gallagher, K. G. & Hawkesworth, C. *Nature* **358**, 57–59 (1992).
- Anderson, D. L. *Theory of the Earth* 1–366 (Blackwell Scientific, Boston, 1989).
- Stein, M. & Hofmann, A. W. *Earth planet. Sci. Lett.* **114**, 193–209 (1992).
- Anderson, D. L. in *Glacial Isostasy, Sea-level and Mantle Rheology* (eds R. Sabadini & K. Lambeck) 379–401 (Kluwer, Dordrecht, 1991).
- Anderson, D. L. *Science* **213**, 82–87 (1981).
- McKenzie, D. *Earth planet. Sci. Lett.* **95**, 53–72 (1989).
- Menzies, M. A. in *Magmaism and the Causes of Continental Break-up* (eds B. Storey, T. Alabaster & R. Pankhurst) 31–39 (Geological Society, London, 1992).

Exaggerated danger

Aaron Wildavsky

Phantom Risk: Scientific Inference and the Law. Edited by Kenneth R. Foster, David E. Bernstein and Peter W. Huber. MIT Press: 1993. Pp. 457. £35.95; \$39.95.

Element of Risk: The Politics of Radon. By Leonard A. Cole. AAAS Press: 1993. Pp. 246. \$29.95.

READING these splendid books together gives the layman (that is, all of us outside our specializations) most of the knowledge we need to understand the claims of harm to human beings from a variety of industrial and natural chemicals. (Atmospheric issues are not covered and work exposures are not treated in depth.) The conclusions of both volumes are (and ought to be) deeply disturbing. In *Phantom Risk*, the editors write: "By *phantom risk* we mean cause-and-effect relationships whose very existence is unproven and perhaps unprovable." How then to resolve the resulting legal claims?

But perhaps they are really unresolvable. The proclaimed causes are so small and the statistical associations so weak that the effects become undecipherable. Many studies run foul of "the multiple comparison effect. An investigator who makes enough comparisons in a study will surely find many 'statistically significant' associations in the data, merely because of chance." Recently, for instance, there were reports from Sweden about harm to children from electromagnetic forces. The trouble was that this harm was not the same kind of harm as reported elsewhere in relation to electromagnetic radiation. In considering such issues we need to go back to Darwin who taught that 'findings' unrelated to expectations generated by theories are worthless.

The claims of harm do not meet most (and often any) of the ordinary canons of scientific evidence, including the Henle-Koch-Evans postulates. Travelling down the postulates, the authors of these 15 studies, who are scientists and lawyers and sometimes both, find that the claimed ills do not appear more often in those exposed than in controls; that in prospective studies, which begin with a presumably healthy population, the exposed do not reveal a higher incidence of diseases; that replication of the experiment does not result in the same effects, and so on.

Last, but not least, the vast bulk of regulation is based upon animal tests using rodents, which, though useful for many purposes, are increasingly falling into scientific disrepute as a means of predicting cancer in human beings. As two contributors to *Phantom Risk* observe, the US Environmental Protection Agency (EPA) "operates on the assumption that all animal carcinogens lead to cancer risk in humans at all levels of exposure."

The huge doses of chemicals fed to the animals make human comparisons suspect. Bruce Ames and Lois Gold argue that the dose is the demon, that the poor creatures are effectively being poisoned, leading to massive cell division and hence to cancer. Like the famous Hawthorne effect, in which the experiment itself caused enhanced worker performance, the test and not the hypothesized chemical cause is the killer.

The most powerful argument against

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

High tension — is there a link between magnetic fields and cancer?

carcinogen testing in rodents in my opinion is statistical (Freedman and Zeisel *Statistical Science* 3, 60–71; 1988). Given the vast differences in size and dose between rodents and humans, statistical models are needed as controls. Without knowledge of the mechanism of cancer causation, however, there is no way of choosing a model except at random. In the end, all one knows is that the correct result can vary in magnitude from 4,000 times to 4,000,000 times. Is that second best good enough? The response to this criticism has been that, because one should not experiment on humans, rodents are the only alternative. This is ridiculous. One could say the same of bloodletting. One should improve knowledge not ignorance.

A few selections from *Phantom Risk* illustrate the problems covered. In 'Weak Magnetic Fields: A Cancer Connection?', Kenneth R. Foster observes that the

Earth's magnetic field has a flux density of some 500 milligauss; below high-voltage transmission lines the flux densities vary from 30 to 100. Asked whether weak fields might be connected to diseases such as childhood leukaemia, Foster, reviewing the evidence, would say that it is possible; asked whether the cancer connection is probable or plausible, he would give a resounding no. Nevertheless, a jury voted and a judge upheld a verdict against an electrical company in a case where pupils at a nearby school were exposed to between 1.5 and 2 milligauss.

Louis Lasagna and Sheila R. Shulman, in 'Bendectin and the Language of Causation' conclude that "the scientific evidence seems sufficient to rule out the possibility that Bendectin is a powerful cause of birth defects. The possibility that it might cause undetectably small increases in the rate of birth defects cannot be ruled out by scientific data. Proving that Bendectin does not cause birth defects is logically impossible." The movement from probabilities to possibilities, and from positive evidence of harm to negative evidence of no harm is important. For although science can be of great help in dealing with probabilities and positive evidence, it is of much less help in establishing impossibilities and 'never-could-bes'.

After establishing the significant differences between different types of asbestos fibres and levels of exposure, Ralph D'Agostino Jr and Richard Wilson observe that children in schoolrooms containing asbestos are less at risk than pedestrians are from being hit by a car. Their concluding sentence, which they say sums up the wisdom of the field, is interesting because it focuses attention not on possible danger but on probable safety: "There is no substitute for chrysotile asbestos [used in buildings] that, if properly applied, is *known* to be as safe."

To turn to the second volume under review, Leonard Cole is so deft in analysing the science and politics of radon that it is hard to know whether to laugh or cry. On the one hand, EPA's public pronouncements vastly overestimate the danger from ionizing radiation in basements by extrapolating directly from huge exposures in mines to immensely lower doses in homes. On the other hand, when EPA goes to Congress it is met with insistent demands not only to reduce permitted levels to what is technically feasible but also, in the language of the 1988 Indoor Radon Abatement Act, that the nation's long-term goal should be to have all its buildings "as free of radon as the ambient air outside of buildings" at a cost of more than \$1,000 billion, were it possible to achieve.

Weighing the evidence carefully, Cole sides with far more modest regulation. Equally valuably, he takes his readers on a tour of radon politics and media treat-

Martin Bond/Science Photo Library

ment. He discovers vast overstatement through the media and in particular a liberal Democratic member's disproportionate interest in ever-more stringent regulation, though he does not (the only fault I could find) enquire critically into why this might be so.

My favourite episode concerns the trip to Sweden and return home of Senator Frank Lautenberg who kept praising Swedish initiative, never realizing that its experts preferred and its laws inscribed much more relaxed levels of regulation (two and a half times less) than in the United States. He (and many others) knew what was good for health even if they did not know what they were talking about.

The ancient molecule

Sydney Brenner

The RNA World. Edited by Raymond F. Gesteland and John F. Atkins. *Cold Spring Harbor Laboratory Press: 1993. Pp. 630. \$95.*

THE characteristic feature of contemporary biological systems is that they use two chemical languages: function is written in the 20 amino-acid 'alphabet' of proteins while information is inscribed in sequences of four nucleotide 'letters' (codons). Translation from the genetic to the functional language is carried out by complicated molecular machinery. First a transcript of the gene is made in RNA and then this RNA is translated into protein using transfer RNAs to convert triplet codons into amino acids. Large RNA-containing ribosomes bring all of the components together and catalyse peptide bond formation. Many viruses have RNA genomes and RNA can be replicated both directly and, in some cases, through a cycle that involves conversion to DNA. In higher organisms with genes interrupted by introns (non-coding DNA) there is again an elaborate machinery for splicing these out of the message.

How did this all arise originally? For a long time it has been clear that the contemporary system could not have been the first but must have replaced an earlier system that had a unitary chemical language to allow catalysis and the replication of genetic information to evolve in one molecule. Thus either proteins came first and there was some way for amino acids to be copied, or the first molecules were nucleic acids, RNA in particular, and these could have had catalytic functions. In 1968, Francis Crick wrote a paper in which he went for the latter alternative, largely because it seemed difficult to create an amino-acid pairing system, and anyway the double helix had completely

Placing these books side by side, we come to an amazing conclusion: most of the claims of harm to human beings from technology are either false or unproven. And when there could be real danger, as with radon, that justifies regulation, the risk is vastly exaggerated. Is it the precautionary principle (when in the slightest doubt, regulate) that has led us astray? Or is it our fascination with minuscule causes and infinitesimal effects that has made science irrelevant and, with science, made us so easy to fool? □

Aaron Wildavsky died on 4 September 1993. He was at the Survey Research Center, 2538 Channing Way, Berkeley, California 94720, USA.

discredited earlier views that genes were made of proteins. What he and others did not suspect was that this ancient world of RNA enzymes had not only left traces of itself in modern systems but was also still alive and in good working order. The discovery by T. Cech and S. Altman of RNA molecules that catalyse cleavage and joining reactions of RNA came as a great surprise and prompted a search for other relics of the RNA world, and attention was immediately directed to many other processes in which the RNA components had previously been thought to play only a structural role.

This dense book, *The RNA World*, brings together all the work in this interesting and exciting field, and for many

scientists, who like myself have followed it only at some distance, it offers a uniquely authoritative account of research on RNA catalysis in biological systems, and theories of how these surviving functions reflect the properties of a prebiotic world of RNA. H. Noller has identified conserved domains of the ribosomal RNAs involved in the many different steps of recognition in translation and has recently shown that the RNA alone can still catalyse a simple analogue of peptide bond formation. If there were transfer-RNA molecules or forerunners of them that could catalyse their own specific amino-acid acylation, then the whole of protein translation could have been derived from an RNA precursor. Many hold the view that the intron-exon structure of genes was present from the very beginning and indeed was a characteristic feature of the RNA world. Cech discovered that there were still self-splicing introns and two chapters, one by M. Moore, C. Query and P. Sharp and the other by S. Baserga and J. Steitz, discuss how small RNA molecules participate in the splicing of RNA transcripts in higher organisms. In passing, it should be noted that this view implies that the genes of the most advanced higher organisms, like ourselves, more faithfully reflect the primitive genome structure than do the genes of contemporary microorganisms, which have become more sophisticated by streamlining and have eliminated their introns.

Molecular biology has provided tools that enable us to simulate evolution experiments with RNA in the laboratory. DNA sequences can be amplified by the polymerase chain reaction and then



HEAVENS above — one view of the Trifid nebula. From *A View of the Universe* by David Malin. Cambridge University Press, £24.95, \$39.95.

copied into RNA, and there also exist systems in which RNA molecules can be amplified through a DNA intermediate by a combination of reverse transcriptase and polymerase. Thus, immense populations can be handled and molecules with various properties can be iteratively selected. Two chapters, by L. Gold *et al.* and by J. Szostak and A. Ellington, discuss this burgeoning field of artificial genetics. Gold has searched RNA sequence space for ligands and he (and others) has discovered RNA molecules that bind not only to proteins that normally have nucleic acid ligands but also to others such as thrombin and nerve growth factor that do not. It still remains to be seen how easy it is to pass from this sequence space to that of the simple heterocyclic molecules favoured by the pharmaceutical industry.

In a coda to his paper, Gold has some stringent remarks to make about scientists (like myself) who view nucleic acids "as strings of information rather than sources of interesting globular ligands". In his enthusiasm to simulate "experiments done billions of years ago" (with nerve growth factor?) he has forgotten that there may also be a need to simulate modern systems as well and his particular objections can be easily dealt with. RNA molecules provide the easiest implementation for exploration in combinatorial chemistry, but there are other technologies which can and will be used to explore test-tube molecular evolution.

Although, in principle, these discoveries with RNA molecules, both natural and artificial, narrow the gap between primitive genetic systems in which function and information were carried and copied within the same molecule and the modern bilingual systems, they do not help much with the origin of life itself. In a sobering first chapter, G. Joyce and L. Orgel point out that there are still enormous difficulties in going from a mixture of purine and pyrimidine bases to any kind of self-replicating system. We seem to have crossed the gap between molecular biology and genetics, but that between chemistry and molecular biology remains wide and deep. In fact, some people have worried that there may not have been enough time to have done it all in our solar system. Crick refers in his foreword to his hypothesis of directed panspermia which gains time by having the early steps going on somewhere else in the Universe followed by the transport to Earth of a population of microorganisms a few billion years ago to continue the good work here. Thus, when we come to the pre-RNA world, we will have to decide not only what to look for but also where to look for it. □

Sydney Brenner is in the Department of Medicine, University of Cambridge School of Clinical Medicine, Addenbrooke's Hospital, Hills Road, Cambridge CB2 2QQ, UK.

Toxic barb

Alastair Hay

General and Applied Toxicology, Edited By Bryan Ballentyne, Timothy Marrs and Paul Turner. Stockton/Macmillan Press: 1993. Pp. 1,456. \$295, £195.

ASSESSING a general text on toxicology is not that far removed from assessing the potential risk of a particular chemical. It is necessary to know what each does. How specific is the available information? Are underlying mechanisms explained and dissected? Further considerations include the treatment of general information about target organs involved, effects on the environment and, it is to be hoped, discussions about species differences.

General and Applied Toxicology meets

IMAGE UNAVAILABLE FOR COPYRIGHT REASONS

The military applications of toxicology — a victim of chemical warfare.

these criteria. Aimed at students of the subject, the book is a mine of information for both the generalist interested in, for example, the principles underlying toxicity testing, and for the specialist concerned about gases released during combustion and the damage they can cause.

The book is intentionally wide-ranging. Well-known researchers have contributed review chapters on the subject of their speciality. Published in two volumes, the book is further subdivided into sections dealing with basic concepts such as study design, haematology and clinical chemistry; toxicity by contact route; target organs; specialist toxicology, which includes chapters on mutagenicity, epigenetic carcinogenesis and neonatal toxicology; regulatory toxicology; and, finally, single issues such as the role of epidemiology, or the toxicology of pesticides, che-

mical warfare agents and food additives.

One particularly pleasing feature of the book is the crispness of the chapters. Most are very well written and concise as a result, presumably, of a firm editorial hand. This conciseness enables the authors to be wide-ranging in their coverage of topics and, for some readers, sections may be all too brief. The chapters are, however, extremely well referenced and more interested specialists will find plenty to guide them to other research sources.

The effects of explosive devices are not addressed specifically, although chapters on combustion toxicology and chemicals used in warfare to asphyxiate or to poison steer the book around the area of the military applications of the subject. For some scientific researchers, explosive devices may have represented all too real a hazard. As targets of animal rights activists and with growing public concern about the welfare of animals, toxicologists have had their share of bad press. The criticisms have also brought rewards. Increasing emphasis is now placed on the use by researchers of well-validated *in vitro* systems of testing.

Such tests have revolutionized the investigation of mutagens and potential carcinogens and may do the same when applied to many other areas of toxicology through the use of cell cultures from specific organs. These may in time be useful predictive models for the effects of chemicals on target organs. *In vitro* systems and the ethical issues that have to be dealt with by toxicologists are also aired in the book.

General and Applied Toxicology tends to emphasize the pragmatic approach to the regulation of chemicals and drugs exemplified by countries in the European Union, and by the Food and Drug Administration in the United States. Chapters on regulation review the processes involved and, according to the editors, regulatory toxicology is becoming "increasingly dominant, though sometimes scientifically suspect". This comment is a barb directed, no doubt, at some regulatory agencies, such as the US Environmental Protection Agency, which use predictive models for estimating cancer risk for substances at very low doses. Such models can vary by as much as 10^6 in their prediction of the risk involved, and are extrapolated from data obtained from high-dosage tests. These models are not used in the United Kingdom for reaching regulatory decisions.

Although far from cheap, *General and Applied Toxicology* helps to fill a particular niche in the toxicology literature. Purchasers are likely to judge that their money has been well spent. □

Alastair Hay is in the Chemical Pathology Department, University of Leeds, Leeds LS2 9JZ, UK.

Unifying science

Knut Schmidt-Nielsen

The Logic of Life: The Challenge of Integrative Physiology. Edited by C. A. R. Boyd and D. Noble. Oxford University Press: 1993. Pp. 226. £8.95, \$14.95.

How can the avalanche of information about molecular biology be put together into a conceptually coherent picture? In their introductory essay to this book, Noble and Boyd provide an answer, they conclude that "there can be no doubt that physiology . . . will be indispensable to the proper interpretation of molecular biology itself".

They compare the feeling we get from the astonishing progress in molecular biology today to the excitement in the middle of the nineteenth century when German organic chemists forged ahead with spectacular discoveries of 'organic' substances in 'non-living' test-tube environments. This analogy implies that the role of molecular biology is in question; they answer that without an understanding of the overall logic of higher-level systems, the lower-level detail becomes mere cataloguing of information.

Noble and Boyd sent their essay to a number of eminent biologists, asking them to respond in whatever way was appropriate to their own field. This intellectual challenge resulted in eight essays that represent the wide spectrum of biological thinking today, yet, in its own way, each illuminates the importance of the organism. With the editors' introductory essay and a foreword by the Nobel laureate Sir James Black, these essays form the chapters of this small volume published in celebration of the 1993 International Congress of Physiology.

The essays are remarkably diverse, which should be no surprise as only 4 of the 12 contributors are associated with physiology departments. In spite of the vast differences in approach, there is a common theme, nearly all deal with the living organism in the broadest sense, how it operates and what makes it tick. To paraphrase what runs through many of the essays, it isn't the genome and its DNA that makes the organism, it is the organism that makes DNA.

This is clear in Stephen J. Gould's discussion of the controversies pertaining to the rate of evolutionary change. He asserts that the triplet code is only machine language (a metaphor for which he thanks E. F. Yates), and that the program resides at a higher level of control and regulation — and that we know virtually nothing about it. Almost hidden in his essay is the important statement that development is the only discipline of biology that might unify molecular and evolutionary approaches into a coherent science.

In a related chapter on development, R. L. Gardner and C. D. Stern write that the central aim of contemporary developmental biology is to explain how a single cell is transformed into a complex multicellular organism. How is development regulated? It is so easy to fall into the trap that it is all programmed in the genome; the DNA is no more able to build a cell, not to mention a whole organism, than a handbook on architecture is able to build me a new house.

Likewise, the question of regulation is implied in Jared Diamond's discussion of safety factors in living organisms. His considerations are based in evolutionary physiology; building more structure than is necessary is not only expensive, its maintenance is a further metabolic cost. The main question then becomes, how

much is enough but not too much? A high-carbohydrate diet causes upregulation of the intestinal glucose transporter and physical training increases muscular and circulatory performance. This again pertains to questions of regulation at a level of organs and organism.

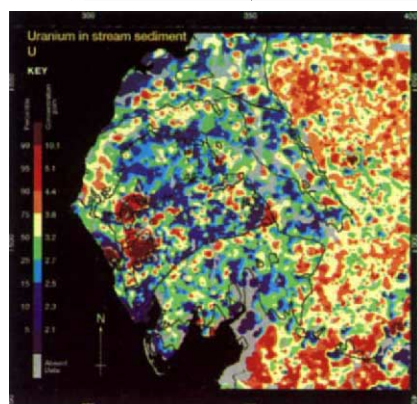
Regulation is a main theme in the chapters by Derek Denton and Jean Didier Vincent. When a sheep is made sodium deficient, it will within a few minutes drink the precise amount of sodium bicarbonate solution to match the deficiency, and then stop long before feedback mechanisms can take effect. The how and why addressed to physiologists include such puzzling questions as why neurohormones are generally released in a pulsatile way. The function significance of pulsatility is unknown and the nature of the pulse-generator remains a mystery. We only know, for example, that shifting from pulsatile to a continuous infusion of gonadotropin-releasing hormone rapidly inhibits gonadotropin secretion.

In a discussion of sensory physiology and brain function, M. Konishi emphasizes the importance of considering behaviour. The brains of owls contain neurons that respond only to sounds coming from a particular direction, and had the investigators not known about the perceptual problems the animal must solve in real life, they would not have looked for neurons selective for specific stimuli. He emphasizes that, to achieve its goal, brain sensory physiology must treat the system as a whole, perhaps the hardest hurdle to cross for neurophysiologists.

In the concluding chapter, "Self-Organizing Systems", F. E. Yates argues that living systems are neither state-determined nor DNA-determined. In quoting from R. C. Lewontin he calls attention to the absurdity of saying that genes have created us body and mind. "It is often said that DNA produces proteins, in fact proteins (enzymes) produce DNA Not only is DNA incapable of making copies of itself, it is incapable of 'making' anything else Without the protein-forming machinery, *nothing* can be made."

I have chosen what I perceive as a consistent message in this book, and I feel comfortable in concluding, as the editors did, that physiology will be indispensable in putting together and interpreting the masses of detailed information emanating from the revolutionary progress in molecular biology. We must detach ourselves from the infatuation with more data and more information and remember that, in the end, all the interesting molecules come from, and belong in, living organisms. □

Knut Schmidt-Nielsen is in the Department of Zoology, Duke University, Durham, North Carolina 27708-0325, USA.



Hot streams — concentrations of uranium in stream sediments in the Lake District of the United Kingdom (left). Right, false-colour composite image of the same area using near and shortwave data acquired by the Landsat Thematic Mapper sensor. These images are taken from the most recent volume in the British Geological Survey series (*Regional Geochemistry of the Lake District and Adjacent Areas*), which aims to provide a systematic picture (in large format) of the geochemistry of the whole of Great Britain starting in the north and working southwards. British Geological Survey, £50.

The Permo–Triassic extinction

Douglas H. Erwin

The end-Permian mass extinction brought the Palaeozoic great experiment in marine life to a close during an interval of intense climatic, tectonic and geochemical change. Improved knowledge of latest Permian faunas, coupled with recent advances in isotopic studies and biostratigraphy, have greatly enhanced our understanding of the events of 250 million years ago and have begun to provide answers to many questions about the causes of extinction.

KILLING over 90% of the species in the oceans^{1,2} and about 70% of vertebrate families on land^{3–5} is remarkably difficult. The end-Permian mass extinction was the closest metazoans have come to being exterminated during the past 600 million years. The effects of this extinction are with us still, for it changed the structure and composition of marine communities far more than any event since the Cambrian radiation. The end-Permian extinction brought the world of the Palaeozoic to a close, permitting the expansion of new marine community types which continue to dominate modern oceans. This event plays a critical role in debates over the nature of mass extinctions and their role in structuring the evolution of life.

By 1840, Phillips⁶ recognized that the history of macroscopic life on Earth could be divided into three great eras, which he named the Palaeozoic, Mesozoic and Cenozoic; he also recognized two dramatic drops in diversity, each associated with the appearance of new types of organisms. Yet until well into this century, palaeontologists paid little attention to these events. Since publication of the impact hypothesis⁷, interest has focused on the Cretaceous/Tertiary (K/T) mass extinctions and other events with a possible extraterrestrial cause. In part this stems from the widespread, if erroneous, view that the record across the Permo–Triassic (P/Tr) boundary is too poor for detailed study, a presumption based on apparent evidence for a widespread marine regression and depositional hiatus at the boundary.

The pattern of disappearances across the Permo–Triassic boundary is complex, with some clades disappearing well below the boundary, others quite diverse right up to the boundary, and still others seemingly oblivious to the extinction^{2,8}. Analysing such patterns is complicated by distortions including the regression, backward smearing of true last occurrences⁹, and many Lazarus taxa. Lazarus taxa¹⁰ disappear from the record during the Late Permian, but did not become extinct, for they reappear in Middle Triassic rocks. First recognized in gastropods¹¹, the phenomenon is widespread among bivalves, brachiopods and

other taxa, demonstrating both the extent of sampling problems and the importance of undiscovered refugia in preserving many lineages.

Detailed palaeontological, geochemical and sedimentological studies of sections in Italy, Pakistan and south China are revealing a rich and complex pattern of events. New correlations based on conodonts permit more precise and more accurate intercontinental correlations and comparison of extinction patterns among regions. These new data have, in turn, led to the re-examination of older hypotheses about the causes of the extinction and the proposal of additional scenarios.

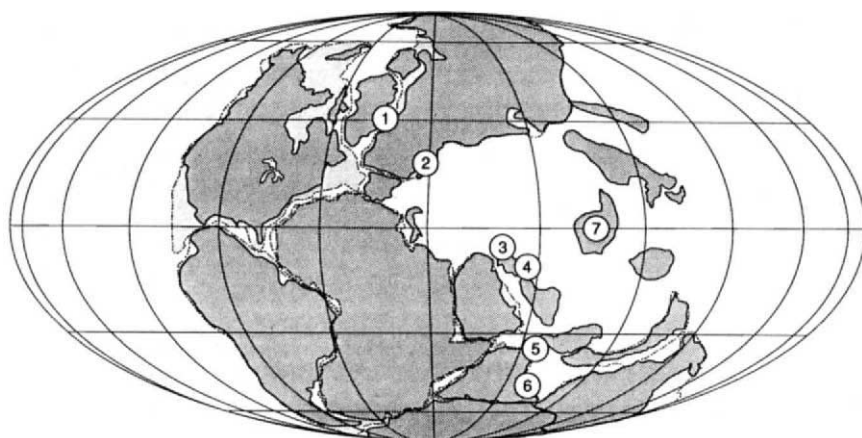
The cause of the end-Permian mass extinction appears to involve a tangled web rather than a single mechanism. Three phases can be identified. The first began with the onset of the marine regression which dried out many marine basins, reduced habitat area and increased climatic variability. The regression accelerated during the second phase, triggering the release of gas hydrates and the erosion and oxidation of marine carbon. In conjunction with the eruption of the Siberian flood basalts, these carbon sources exacerbated climatic instability; an increase in atmospheric carbon dioxide may have produced oceanic anoxia and global warming. The final phase of the extinction involved the destruction of near-shore terrestrial habitats during the rapid earliest Triassic marine transgression.

Marine extinctions

The marine fossil record provides the most complete and detailed record of the extinction. Examining the temporal, biogeographical and ecological pattern of extinction allows evaluation of proposed extinction scenarios, all of which ultimately requires high-resolution biostratigraphic correlation between the sections spanning the Permo–Triassic boundary.

Boundary sections. The Permo–Triassic boundary was traditionally recognized by the first occurrence of the ammonoid *Otoceras woodwardi* and the more easily identifiable bivalve *Claraia*. Correlating between the primary boundary sections

FIG. 1 Continental positions in the Late Permian, showing the location of important Permo–Triassic boundary sections. The number and continuity of individual sections in the southern Alps, Iran and south China belies earlier claims for a worldwide period of non-deposition spanning the Permo–Triassic boundary. (1) Greenland; (2) Southern Alps; (3) Iran–Armenia border (Kuh-e-Ali Bashi section); (4) Central Iran; (5) Salt Range, Pakistan; (6) Guryl Ravine, Kashmir; (7) numerous sections in South China.



(Fig. 1) has been difficult, but correlations based on conodonts have resolved these problems, although they have dramatically altered the relationships among boundary sections (reviewed in refs 2, 8, 12 and 13, but see ref. 14) (Fig. 2). These results suggest *O. woodwardi* first appeared during latest Permian boreal, cold-water faunas¹³. In contrast, *Otoceras* is absent from contemporary warm-water faunas of the Tethyan realm, including the magnificent Changxingian faunas of South China. Moreover, these correlations offer the prospect of the first global, high-resolution analysis of extinction and survival patterns.

Contrary to earlier claims¹⁵, recent studies of the Permian Bellerophon Formation and the overlying Werfen Formation in the Alps indicate that this classic sequence preserves a detailed palaeontological¹⁶ and geochemical¹⁷ record of the extinction interval. The boundary lies within the Tesero Horizon at the base of the Werfen Formation, which contains a fauna transitional between the Permian and Triassic, similar to mixed faunas from South China¹⁸. A gradual shift in carbon isotopes ($\delta^{13}\text{C}$) from +3‰ to -1‰ is associated with the extinction horizon¹⁷. A curious spike in spores of the fungus *Tympanocysta*¹⁹ may indicate widespread collapse and decay of terrestrial ecosystems.

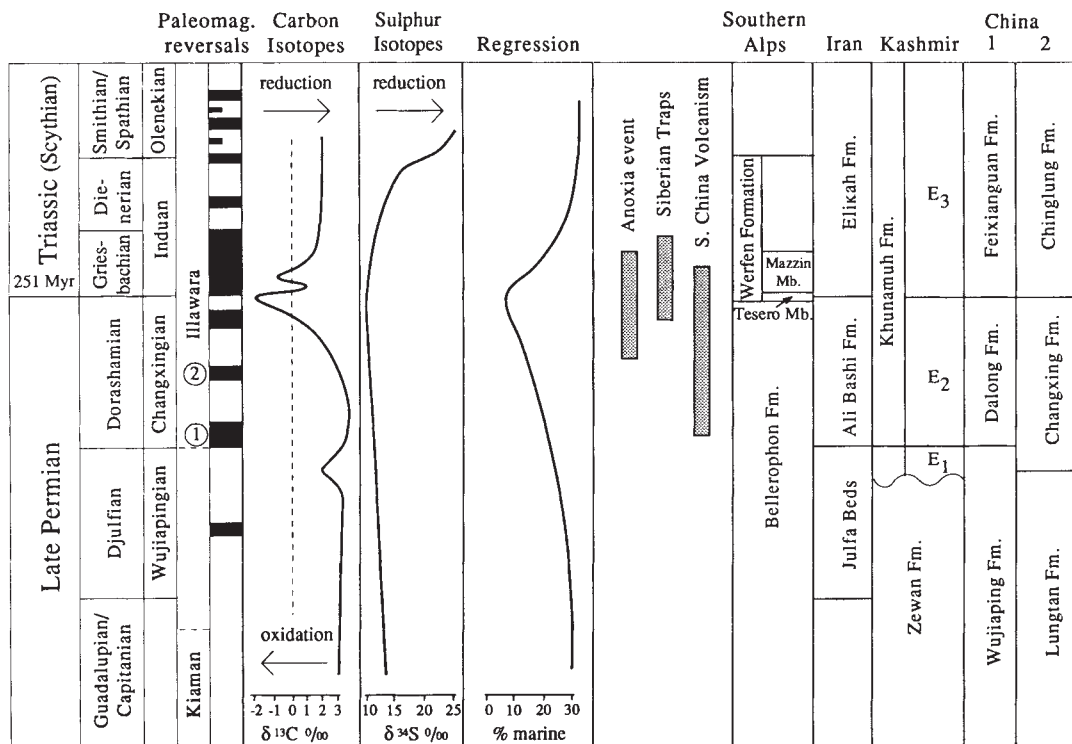
The widespread Permo-Triassic boundary deposits in South China²⁰⁻²⁴ contain a diverse fauna. In the latest Permian Changxing Formation, 435 of 476 invertebrate species, or 91%, disappear (ref. 24, cited in ref. 25), including 98% of ammonoids species, 85% of bivalve species and 75% of the shallow-water fusulinid foraminifera. However, taxonomic inconsistencies and lack of Lower Triassic sediments may overstate the magnitude of the extinction. Furthermore, in contrast to the gradual shift in $\delta^{13}\text{C}$ in the Alps, sections of South China record a relatively abrupt shift²⁶, suggesting that the Chinese sections are condensed relative to those in the Alps. This may account for claims of a catastrophic extinction in the Chinese sections²⁷. The youngest known Palaeozoic reefs are found in the Changxing Formation,

in Sichuan and Hubei Provinces²⁸. Thus a diverse Permian reef fauna disappears before the boundary as evaporites develop, but their presence belies claims of prolonged environmental deterioration before the P/Tr boundary, at least in South China.

Clay layers, apparently of volcanic origin, occur throughout the latest Permian Changxing formation and a particularly thick clay marks the basal Triassic. A tuffaceous texture, bipyramidal quartz, volcanic shards and geochemistry all indicate an explosive silica-rich volcanic source for the ash²⁹⁻³¹, probably associated with a subduction zone. This ash covers $>10^6 \text{ km}^2$ (ref. 30), representing eruption of 1,000–4,000 km^3 of material³¹. This volcanism has been linked to the extinction^{20,24,29,30}, but is no larger than many other pyroclastic eruptions, with no discernable biological effects³². Reports of iridium enrichment within boundary clays^{25,33,34} cannot be replicated^{30,35,36}, arguing against the kind of impact associated with the Cretaceous-Tertiary mass extinction. The lack of an impact signature also raises questions about the nature of the cause of the apparently periodic mass extinctions stretching from the late Permian into the Miocene³⁷. Immediately overlying the boundary clay are a series of characteristic transitional beds containing three distinctive mixed faunas of both Permian and Triassic fossils²³; remixing after deposition has been ruled out as a cause of this assemblage³⁸.

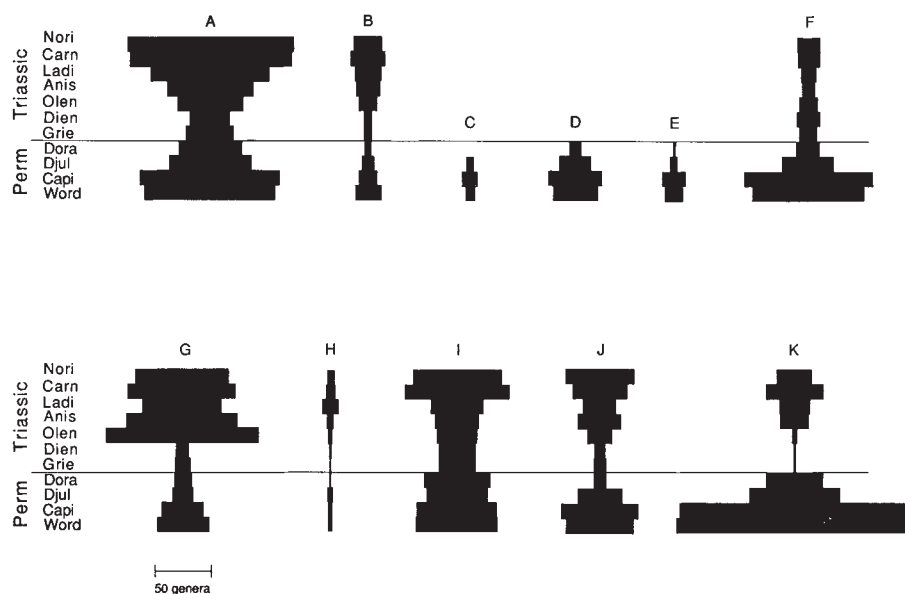
Extinction patterns. Global diversity of durably skeletonized marine families declined from 536 to 267 (49%) between the Capitanian and the end of the Permian; about 72% of corresponding genera disappeared^{37,39-41}. In comparison, only 57% of genera disappeared during the end-Ordovician mass extinction, the second largest of the Phanerozoic³⁷. Rarefaction analysis suggests that global species extinction was $>90\%$ (ref. 1), although the magnitude and pattern varied considerably among different clades (Figs 3 and 4). This included a 79% extinction among families from Sepkoski's Palaeozoic evolutionary fauna (predominantly epifaunal, suspension feeders, including Palaeo-

FIG. 2 Permo-Triassic geological events and correlation between major boundary sections. Standard geological stages are shown to left, with two alternative names given for latest Permian. Zircons from the boundary clay at Meishan, China provide a date for the Permo-Triassic boundary of $251.2 \pm 3.4 \text{ Myr}$ ago⁷². The end of the Kiaman Reversed Superchron and the onset of the mixed polarity interval at the base of the Illawara Superchron is not well delineated. It has been placed near the base of the Changxingian stage^{73,74}, but other work suggests it may have occurred as early as the latest Capitanian (M. Menning, personal communication). Black is normal magnetic polarity, white is reversed polarity (1, *Neogondolella orientalis* zone; 2) base of the *N. changxingensis* zone. Shifts in carbon isotopes¹⁷ and sulphur isotopes⁵⁰ are schematic. Strontium isotopes show a similar pattern to the carbon isotopes, declining to the lowest value of the Phanerozoic near the boundary, indicating an influx of light strontium from juvenile basalts, generally on mid-ocean ridges. Other significant events include the eruption of the massive Siberian flood basalts, a lengthy (1–3 Myr)



anoxia event in deep-water marine rocks in Japan (Y. Isozaki, personal communication), a widespread pyroclastic volcanism in south China. Correlation between boundary sections in the southern Alps; Kuh-e-Ali Bashi, northeastern Iran; Guryl Ravine, Kashmir; and Shangsi (1) and Meishan (2) in southern China are based on conodont analyses¹².

FIG. 3 Generic diversity patterns among marine invertebrates from the mid-Permian through to the Triassic. All stages are of equal width, although they range in duration from 1–2 Myr for the Djulfian, Dorashamian and Griesbachian to ~13 Myr for the Norian. There is great variability in extinction–survival–recovery pattern. Nautiloids (B), sponges (I), non-fusulinid foraminifera, conodonts and some gastropods (none shown) were almost oblivious to the extinction; although some species disappeared, the level of extinction was far less than among other taxa. A second group appears to decline together with the marine regression. These include the tabulate and rugose corals (C, D), the trilobites (E), the bryozoa (F), the articulate brachiopods (J, K) and the crinoids and blastoids, although the poor fossil record of the last two during the Permian makes this pattern suspect. This pattern is partly an artefact of the declining quality of the fossil record, however, because many of these taxa, including fusulinid foraminifera, articulate brachiopods, ammonoids (G) and several gastropod groups exhibit high diversity in latest Permian rocks, particularly in South China, but decline rapidly near the boundary. Abbreviations: Word, Wordian; Capi, Capitanian; Djul, Djulfian; Dora, Dorashamian (=Changxingian); Grie, Griesbachian; Olen, Olenkian; Anis, Anisian; Ladi, Ladinian; Carn, Carnian; Nori, Norian. A, Bivalvia; B, Nautiloidea; C, Tabulata; D, Rugosa; E, Trilobita; F, Bryozoa; G, Ammonoidea; H, Echinodermata; I, Porifera; J, Rhynchonellida + Terebratulida (brachiopods);



K, all other articulate brachiopods; mostly spiriferids in the Triassic. These data have not been corrected for the latest conodont correlations and suffer from the deficiencies of such broad scale-compensia. Nonetheless, they provide the best available overview of extinction and survival patterns. The new conodont correlations¹² will permit higher-resolution analysis of the immediate boundary interval. Data provided by J. J. Sepkoski Jr.

zoic corals, articulate brachiopods, stenolaemate bryozoans and stalked echinoderms). In contrast, the modern evolutionary fauna (gastropods, bivalves, echinoids) declined by only 27% (ref. 41).

The reliability of the patterns shown in Fig. 3 is unclear. For example, the decline of the foraminiferid suborder Fusulinina was thought to have begun by mid-Permian times. Yet fusulinids are diverse in the Changxing Formation before disappearing rapidly near the boundary. Non-fusulinid forams, which lived in deeper water, experienced only slight extinction, although architecturally more complex taxa seem to suffer more than less complex forms across all suborders⁴². Does this mean the apparently gradual extinction simply reflected the loss of shallow marine sediments outside of South China? Perhaps, but South China was an isolated tectonic block during the latest Permian (Fig. 1). It may have served as a refugium for groups already declining on Pangea. Palaeontologists do not yet have the high-resolution data on global extinction and survival patterns that will resolve these questions. There is, however, little support for claims that the mass extinction occurred over eight million years⁸, but it is unclear whether the extinction lasted two million years, 1 million years or even less.

Several generalities seem secure however. Clades of sessile, epifaunal, filter-feeders generally suffered, but such clades often share other risk factors, including near-shore distribution, or restricted environmental distribution⁴³. The apparent increased extinction among marine invertebrates with planktotrophic larval development^{44,45} may simply reflect the increased extinction in near-shore environments and in the tropics, both areas where planktotrophs dominate.

The terrestrial record

Most palaeontologists have emphasized the extensive marine extinctions and hence most descriptions have emphasized marine processes. Data from terrestrial vertebrates and insects reveal widespread extinctions during the Late Permian, although the data are not sufficiently detailed to resolve the duration or timing of the extinctions.

Although plagued by taxonomic and sampling problems, 21 of 27 families of reptiles and six of nine amphibian families disappeared during the latest Permian, for an overall 75% drop in diversity³. The well studied vertebrate faunas from the Karoo Basin in South Africa reveal two extinction peaks, one roughly correlative with the end of the Capitanian, the other in the upper Permian, but below the Permo-Triassic boundary⁴. This may reflect nothing more than sampling bias and the difficulties correlating within the Karoo Basin and between marine and terrestrial sections, but does suggest a complex extinction event.

Among insects, 27 orders have been recorded from the Permian, of which eight disappear during the Late Permian, four suffer considerable declines in diversity but recover, and three straggle into the Triassic with such reduced diversity that they became extinct during the period⁴⁶. The end-Permian mass extinction induced the most profound changes in insect diversity patterns in the history of the class.

The plant record is far more equivocal. The distinctive plant floras of the Carboniferous are replaced by the Mesophytic flora throughout the Permian with the onset of global warming following the Permo-Carboniferous glaciation, but plant fossils show little direct evidence of mass extinction^{47,48}. Pollen undergoes a marked change at the Permo-Triassic boundary. In addition to the spike in fungal spores noted earlier, gymnosperm pollen virtually disappears and a new, arid-resistant pollen type appears^{2,19,49}. Some of the changes may actually reflect changes in terrestrial communities that accompany the rapid earliest Triassic marine transgression².

Changes in the physical environment

The supercontinent of Pangea formed by the Early Permian, but the end-Permian regression led to exposure of the continent⁵⁰, widespread evaporite deposition⁵⁰, global warming and increased climatic instability⁵¹. Other events include eruption of the Siberian flood basalts (1.5 million km²; the largest flood basalt of the Phanerozoic) in less than 1 million years⁵², evidence of global warming and marked shifts in carbon, oxygen, sulphur and strontium isotopes^{17,49,53} (Fig. 2).

The shifts in carbon isotopes have been linked to the extinction through erosion and oxidation of organic carbon previously sequestered on the continental shelves⁵⁷, or oxidation of deep-sea sapropel-like deposits following overturn of a previously stratified, anoxic ocean^{54, 56}. Each hypothesis postulates oxidation reducing atmospheric oxygen and increasing atmospheric carbon dioxide, perhaps leading to anoxia and global warming, although there are no data on the extent of likely anoxia or warming. Alternatively, the signal may represent the spread of anoxic bottom waters across the shelves during the earliest Triassic transgression^{14, 57, 58}. The geochemical data present difficulties for each of these hypotheses^{2, 59}. In particular, they ignore the impact of methane gas hydrates ($\delta^{13}\text{C} \approx -65\%$) released during the regression². Regressions release hydrates locked in the outer continental shelf^{60, 61}, and may modulate ice ages via negative feedback⁶¹. No negative feedback occurs during non glacio-eustatic regressions and larger volumes of methane can be released.

Extinction mechanisms

Supernovas, declining numbers of marine provinces and salinity changes are among the proposed causes; these are discussed elsewhere^{2, 10, 50, 62, 63}. Recent models emphasize volcanism, extraterrestrial impact and global anoxia, but evaluating these is difficult because they often fail to make specific, unique predictions about extinction patterns. Without such predictions, most scenarios are just-so stories: comforting perhaps, but of little use. Given the lack of consistent patterns of extinction or survival which point to a specific cause, geological evidence is crucial in evaluating these scenarios.

The widespread pyroclastic volcanism in South China and the eruption of the Siberian flood basalts have been invoked as causes of the extinction through global cooling^{20, 24, 29, 30, 50}, although the pyroclastic volcanism is too small to have been effective³². Whereas the Siberian traps apparently began erupting near the boundary, most of the flood basalt was emplaced during the earliest Triassic. Moreover, modelling results suggest that the climatic effects of SO_2 are self-limiting: at large volumes the molecules condense into larger particles⁶⁴. Thus it is unclear if this eruption could have caused sufficient cooling to cause the extinction. The claimed iridium enrichment at the boundary in South China^{25, 33, 34} is not supported^{17, 35, 36, 59}. It has been suggested that a period of bipolar glaciation triggered the extinction⁶⁵, but without tying it to a specific cause. However, this glaciation was mid-Permian, representing the final pulse of the Permo-Carboniferous glaciation (J. C. Crowell, personal communication).

Ignoring the lack of evidence, is the pattern of extinction consistent with global cooling, either from volcanism, impact or glaciation? Not really. Although filter-feeders suffer severely, carbon isotopes provide no evidence for elimination of primary productivity, unlike the K/T boundary⁵⁹. This also fails to explain the differential extinction of fusulinid over non-fusulinid foraminifera. The apparently heightened extinction among shallow-water taxa is more consistent with global warming or habitat destruction than with global cooling.

Perhaps the various anoxia hypotheses provide the answer^{14, 54, 58}. Suggestive evidence has been developed for the spread of anoxic waters associ-

ated with the rapid earliest Triassic transgression^{14, 57, 58}. The geochemical data advanced to support the hypothesis⁵⁷ actually provides little support²; indeed, the shift in $\delta^{13}\text{C}$ permits only a moderate shift in atmospheric oxygen levels, limiting the extent of marine anoxia². The sedimentological data are more promising, but remain subject to alternative interpretations. Additionally, global anoxia is a diversity-independent mode of extinction, thus survival should be enhanced by broad oxygen tolerance and large population size, yet many of the surviving taxa had small population sizes. Additionally, although molluscs contain many groups that are well adapted to dysaerobic environments and had higher-than-average survival, a closer look reveals no association between survival and oxygen tolerance.

Although there is much about this event palaeontologists do

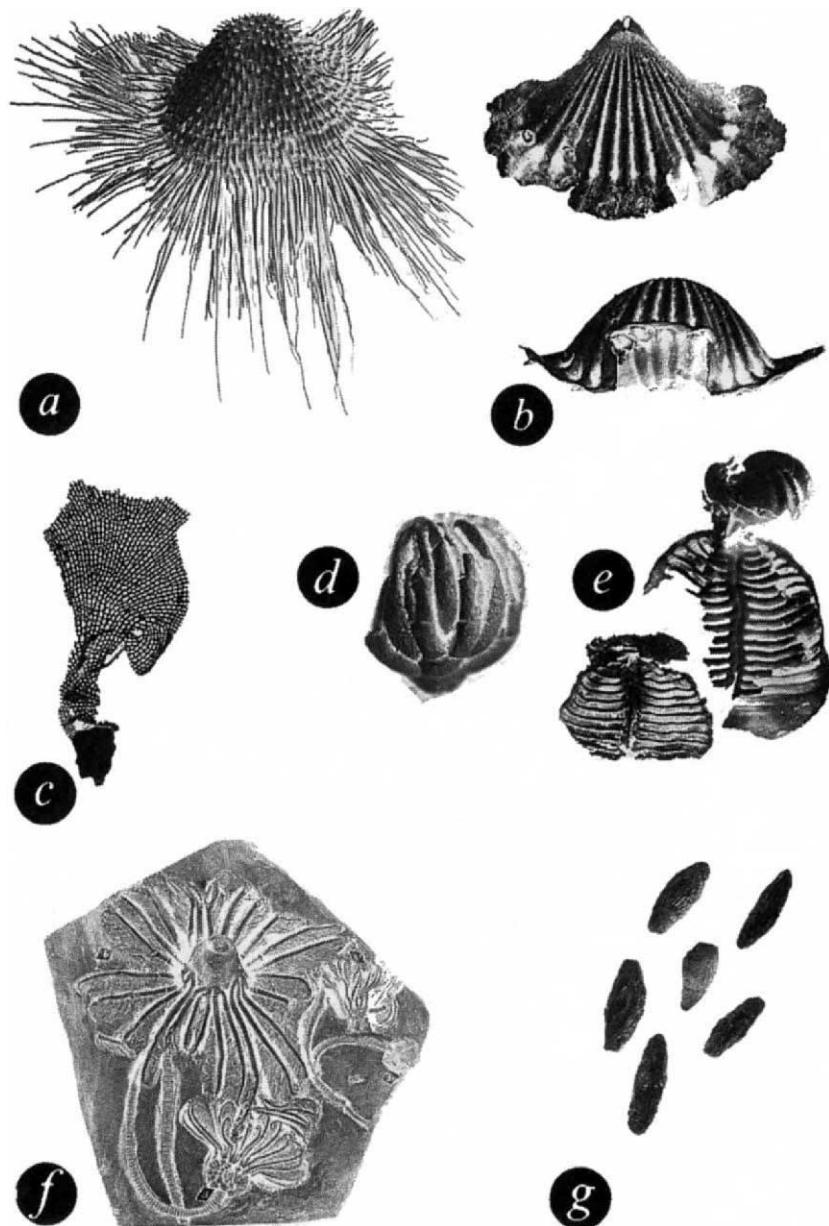


FIG. 4 Representative late Palaeozoic marine invertebrates. Brachiopods a, *Waagenochoncha albichi*; $\times 1.0$, Khisor Range, Pakistan. Late Permian. b, *Stenosisma venustum*; $\times 1.0$, Texas. Lower Permian. c, The bryozoan *Fenestrella* sp.; $\times 0.9$. d, An unusual crinoid, *Timorocrinus multicostatus*, $\times 1.0$, Timor. ?mid-Permian. e, An unusual articulate brachiopod, *Leptotus nobilis*, $\times 0.5$, Khisor Range, Pakistan. f, The Mississippian crinoid *Platycrinus aggaszi*; $\times 0.4$. g, A typical Permian fusulinid foraminifera, *Fusulina* sp.; $\times 1$. Photos a, b, e by R. E. Grant. Photos c, d, f, g by D.H.E. and E. Valiulis.

BOX 1 Life and times of *Miocidaris*

ALTHOUGH echinoids are ubiquitous and familiar members of modern marine communities, only six echinoid genera are known from the Permian and the group might easily have disappeared like blastoids, or other unusual Palaeozoic echinoderm groups. Only a single genus, *Miocidaris* (Fig. 5), is known to have survived the extinction (although cladistic analysis suggests that a related, as-yet undiscovered form survived as well and gave rise to the euechinoids⁶⁶). Its persistence ensured the survival of the class, but fortuitously turned echinoid evolution in a new direction. *Miocidaris* has only two columns of interambulacral plates between each of the five files of ambulacral plates (which cover the tube feet), in contrast to the highly variable number (1–8) of interambulacral plates of other Palaeozoic echinoids. The survival of *Miocidaris* fixed this relationship for all later echinoids.

Was the survival of *Miocidaris* truly fortuitous? Or did this morphological innovation aid the survival of the genus? Would echinoids with two columns of interambulacral plates dominate modern oceans even if the end-Permian extinction had not occurred? These questions are at the centre of one of the great questions in the history of life. To what extent are the long-term patterns in the history of life driven by long-term patterns of adaptation versus the seemingly random patterns of survival during occasional mass extinction events^{43,67,68}? Both equilibrium and non-equilibrium diversity models suggest that the Modern evolutionary fauna would have displaced the Palaeozoic evolutionary fauna even without help from the Permian mass extinction^{37,40,69}. In this view the lower rates of extinction in the clades comprising the Modern fauna ensured their success. Palaeoenvironmental analysis suggests that elements of the Modern fauna began to displace the Palaeozoic fauna by the Late Ordovician^{70,71}. My analysis of Palaeozoic gastropods (a major component of the modern fauna) indicates the pattern is more complex^{43,68}. Palaeozoic and post-Palaeozoic gastropods are very different, and there is no indication that the post-Palaeozoic groups were expanding before the extinction. For gastropods and echinoids at least, the end-Permian mass extinction was a very critical event in their history.

those most tightly integrated into the dominant community types. Some degree of oceanic anoxia may have developed also but, as described earlier, it does not appear that either global warming or anoxia were sufficient to cause such a massive extinction. The final phase of the extinction occurred in the earliest Triassic. The rapid transgression destroyed near-shore terrestrial habitats, causing the shifts in spores and pollen and perhaps much of the decline in insects and tetrapods.

Prospects

Recent geochemical, biostratigraphic and palaeontological studies have considerably sharpened our understanding of the biodiversity crisis that brought the Permian to a close, and the broad spectrum of environmental perturbations that accompanied it. The development of new correlations brings the promise of high-resolution data on the rate and timing of extinction and survival. Comparative analysis of the palaeoecological, geographical and environmental characteristics of these taxa should in turn provide important clues to survival during the greatest biodiversity crisis in the history of life. □

Douglas H. Erwin is at the Department of Paleobiology, National Museum of Natural History, Washington DC 20560, USA.

1. Raup, D. M. *Science* **206**, 217–218 (1979).
2. Erwin, D. H. *The Great Paleozoic Crisis: Life and Death in the Permian* (Columbia, New York, 1993).
3. Maxwell, W. D. *Palaeontology* **35**, 571–584 (1992).
4. King, G. M. *Hist. Biol.* **5**, 239–255 (1991).
5. Benton, M. J. in *Extinction and Survival in the Fossil Record* (ed. Larwood, G. P.) 269–294 (Clarendon, Oxford, 1988).
6. Phillips, J. *Penny Cyclopaedia* **17**, 153–154 (1840).
7. Alvarez, L. W., Alvarez, W., Asaro, F. & Mitchell, H. V. *Science* **208**, 1094–1098 (1980).
8. Teichert, C. in *Extinction Events in Earth History* 199–238 (Springer, Berlin, 1990).
9. Signor, P. W. & Lipps, J. H. *Geol. Soc. Amer. Sp. Pap.* **190**, 291–296 (1982).
10. Jablonski, D. in *Dynamics of Extinction* (ed. Elliott, D. K.) 183–299 (Wiley, New York, 1986).
11. Batten, R. L. in *The Permian and Triassic Systems and their Mutual Boundary Mem.* **2**, 669–677 (Can. Soc. Petrol. Geol., Calgary, 1973).
12. Sweet, W. C. in *Permo-Triassic Boundary Events in the Eastern Tethys* (eds Sweet, W. C., Yang, Z. Y., Dickinson, J. M. & Yin, H. F.) 120–133 (Cambridge Univ. Press, UK, 1992).
13. Kozur, H. in *Occ. Publ. Earth Sci. Res. Inst. N. S. 9-b*, 139–154 (Columbia, South Carolina, 1993).
14. Wignall, P. B. & Hallam, A. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **102**, 215–237 (1993).
15. Assereto, R., Bosellini, A., Fantini-Sestini, N. & Sweet, W. C. in *The Permian and Triassic Systems and their Mutual Boundary Mem.* **2**, 176–199 (Can. Soc. Petrol. Geol., Calgary, 1973).
16. Cassinis, G. (ed.) *Mem. Soc. Geol. Ital.* **34**, 1–366 (1988).
17. Holser, W. T., Schönlaub, H. P., Boeckelmann, K. & Magaritz, M. *Abh. Geol. Bund.* **45**, 213–232 (1991).
18. Broglio Loriga, C., Neri, C., Pasini, M. & Posenato, R. *Mem. Soc. Geol. Ital.* **34**, 5–44 (1988).
19. Visscher, H. & Brugman, W. A. *Mem. Soc. Geol. Ital.* **34**, 121–128 (1988).
20. Yang, Z. Y. et al. *Min. Geol. Min. Res., Geol. Mem.*, ser. 2, **6**, 1–379 (1987).
21. Yang, Z. Y. & Li, Z. S. in *Permo-Triassic Boundary Events in the Eastern Tethys* (eds Sweet, W. C., Yang, Z. Y., Dickinson, J. M. & Yin, H. F.) 9–20 (Cambridge Univ. Press, UK, 1992).
22. Zhao, J. K. et al. *Bull. Nanjing Inst. Geol. Palaeo. Acad. Sinica* **2**, 1–112 (1981).
23. Sheng, J. Z. et al. *J. Fac. Sci. Hokkaido Univ., Ser. IV* **21**, 133–188 (1984).
24. Yin, H. F., Xu, G. R. & Ding, M. H. *Sci. Pap. Geol. Internat. Exch.* (27th Intl Geol. Congr.) 195–204 (1984).
25. Xu, D. Y. et al. *Astrogeological Events in China* (Van Nostrand Reinhold, New York, 1989).
26. Baud, A., Magaritz, M. & Holser, W. T. *Geologische Rundschau* **78**, 649–677 (1989).
27. Xu, G. R. *J. China Univ. Geosci.* **2**, 36–46 (1991).
28. Reinhardt, J. W. *Facies* **18**, 231–288 (1988).
29. Yin, H. F. et al. *Acta geologica sinica* **2**, 417–431 (1989).
30. Yin, H. F. et al. in *Permo-Triassic Boundary Events in the Eastern Tethys* (eds Sweet, W. C., Yang, Z. Y., Dickinson, J. M. & Yin, H. F.) 146–157 (Cambridge, Univ. Press, UK, 1992).
31. Zhao, L. & Kyte, F. T. *Earth planet. Sci. Lett.* **90**, 411–421 (1988).
32. Erwin, D. H. & Vogel, T. A. *Geophys. Res. Lett.* **19**, 893–896 (1992).
33. Sun, Y. Y. et al. in *Developments in Geosciences* (ed. Tu, G.) 235–245 (Science Press, Beijing, 1984).
34. Xu, D. Y. et al. *Nature* **314**, 154–156 (1985).
35. Clark, D. L., Wang, C. Y., Orth, C. J. & Gilmore, J. S. *Science* **233**, 984–986 (1986).
36. Yin, H. F. *Newsletter on Strat.* **15**, 13–27 (1985).
37. Sepkoski, J. J. Jr *Jr geol. Soc. Lond.* **146**, 7–19 (1989).
38. Yin, H. F. *Newsletter on Strat.* **15**, 13–27 (1985).
39. Sepkoski, J. J. Jr *Milwaukee Pub. Mus. Contr. Biol. Geol.* **83**, 1–155 (1992).
40. Sepkoski, J. J. Jr in *Patterns and Processes in the History of Life* (eds Raup, D. M. & Jablonski, D.) 277–295 (Springer, Berlin, 1986).
41. Sepkoski, J. J. Jr *Paleobiology* **10**, 246–267 (1984).
42. Brasier, M. D. in *Extinction and Survival in the Fossil Record* (ed. Larwood, G. P.) 37–64 (Clarendon, Oxford, 1988).
43. Erwin, D. H. *Palaios* **4**, 424–438 (1989).
44. Valentine, J. W. *Bull. mar. Sci.* **39**, 607–615 (1986).
45. Valentine, J. W. & Jablonski, D. *Proc. natn. Acad. Sci. U.S.A.* **83**, 6912–6914 (1986).
46. Labandeira, C. C. & Sepkoski, J. J. Jr *Science* **261**, 310–315 (1993).
47. DiMichele, W. A. & Hook, R. W. in *Evolutionary Paleocology of Terrestrial Plants and Animals* (eds Behrensmeyer, A. K. et al.) 205–325 (Chicago Univ. Press, Chicago, 1992).
48. Knoll, A. H. in *Extinctions* (ed. Nitecki, M. H.) 23–68 (Chicago, Univ. Press, Chicago, 1984).
49. Eshet, Y. in *Permo-Triassic Boundary Events in the Eastern Tethys* (eds Sweet, W. C., Yang, Z. Y., Dickinson, J. M. & Yin, H. F.) 134–145 (Cambridge Univ. Press, UK, 1992).



FIG. 5 *Miocidaris* sp. $\times 1.0$ Texas. A representative of the only clade of echinoids (2 species) to survive the end-Permian mass extinction. Note the arrangement of 2 rows of ambulacral plates separated by 2 rows of inter-ambulacral plates. This pattern is found only in this lineage in the Permian. The survival of this clade insured that all subsequent echinoids share this pattern of construction rather than the variable number of rows of interambulacral plates found in other Permian echinoids.

50. Holser, W. T. & Magaritz, M. *Mod. Geol.* **11**, 155–180 (1987).
51. Parrish, J. M., Parrish, J. T. & Ziegler, A. M. in *The Ecology and Biology of Mammal-like Reptiles* (eds Hotton, N., MacLean, P. D., Roth, J. J. & Roth, E. C.) 109–131 (Smithsonian, Washington DC, 1986).
52. Campbell, I. H., Czamanski, G. K., Fedorenko, V. A., Hill, R. I. & Stepanov, V. *Science* **258**, 1760–1763 (1992).
53. Magaritz, M., Krishnamurthy, R. V. & Holser, W. T. *Am. J. Sci.* **292**, 727–739 (1992).
54. Gruszczynski, M., Halas, S., Hoffman, A. & Malkowski, K. *Nature* **337**, 64–68 (1989).
55. Hoffman, A., Gruszczynski, M. & Malkowski, K. *Mod. Geol.* **14**, 211–221 (1990).
56. Malkowski, K., Gruszczynski, M., Hoffman, A. & Halas, S. *Hist. Biol.* **2**, 289–309 (1989).
57. Hallam, A. *Hist. Biol.* **5**, 257–262 (1991).
58. Wignall, P. B. & Hallam, A. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **93**, 21–46 (1992).
59. Holser, W. T. & Magaritz, M. *Geochim. cosmochim. Acta* **56**, 3297–3309 (1992).
60. Kvernolden, K. A. *Chem. Geol.* **71**, 41–51 (1988).
61. Nisbet, E. G. *Can. J. Earth Sci.* **27**, 148–157 (1990).
62. Erwin, D. H. A. *Rev. Ecol. Syst.* **21**, 69–91 (1990).
63. Maxwell, W. D. in *Mass Extinction: Processes and Evidence* (ed. Donovan, S. K.) 152–173 (Belknap, London, 1989).
64. Pinto, J. P., Turco, R. P. & Toon, O. B. *J. geophys. Res.* **94**, 11165–11174 (1989).
65. Stanley, S. M. *Am. J. Sci.* **288**, 334–352 (1988).
66. Smith, A. B. & Hollingsworth, N. T. *J. Proc. York. Geol. Soc.* **48**, 47–60 (1990).
67. Jablonski, D. *Phil. Trans. R. Soc. B* **325**, 357–368 (1989).
68. Erwin, D. H. *Paleobiology* **16**, 187–203 (1990).
69. Kitchell, J. A. & Carr, T. R. in *Phanerozoic Diversity Patterns* (ed. Valentine, J. W.) 277–309 (Princeton Univ. Press, Princeton, 1985).
70. Sepkoski, J. J. Jr & Miller, A. I. in *Phanerozoic Diversity Patterns* (ed. Valentine, J. W.) 153–189 (Princeton Univ. Press, Princeton, 1985).
71. Sepkoski, J. J. Jr *Paleobiology* **17**, 58–77 (1991).
72. Claoue-Long, J. C., Shang, Z. C., Ma, G. G. & Du, S. H. *Earth planet. Sci. Lett.* **105**, 182–190 (1991).
73. Nikistivskiy, E. A. *Int'l Geol. Rev.* **34**, 1001–1007 (1992).
74. Steiner, M., Ogg, J., Zhang, Z. & Sun, S. *J. geophys. Res.* **94**, 7343–7363 (1989).

ACKNOWLEDGEMENTS. I thank W. T. Holser and K. Towe for discussion, D. Jablonski and J. J. Sepkoski Jr for reviews, and E. Vaiulis for drafting the figures. This work was supported by the Charles D. Walcott Fund of the Smithsonian Institution.

ARTICLES

Regulation of progression through the G1 phase of the cell cycle by the *rum1*⁺ gene

Sergio Moreno* & Paul Nurse

ICRF Cell Cycle Group, Department of Biochemistry, Oxford University, Oxford OX1 1QU, and ICRF, PO Box 123, Lincoln's Inn Fields, London WC2A 3PX, UK

The *rum1*⁺ gene is identified as a new regulator of G1 progression in fission yeast. It influences three aspects of G1 regulation: determination of the length of G1, dependence of S phase upon completion of mitosis, and restraint of mitosis until G1 is finished. We propose that it has a central role in regulating the G1 phase of the cell cycle.

THE major controls of the eukaryotic cell cycle are concerned with regulating the onset of S-phase and mitosis, and with ensuring that these two events are coupled together in the correct order. In the fission yeast *Schizosaccharomyces pombe*, two requirements are important for determining the onset of S phase. The first is the need for a cell to attain a minimum critical cell mass^{1–3}. Small cells in G1 cannot undergo S phase until they have attained this mass, and the time taken to reach this mass determines the length of G1. The second is the need for a cell to complete mitosis, demonstrated by the fact that cells blocked in G2 cannot undergo S phase^{3,4}. In both budding and fission yeast these two requirements are thought to operate through Start, the point in G1 where the cell becomes committed to the cell cycle^{1,5–7}. The molecular basis for Start is not fully understood, but in fission yeast involves the p34^{cdc2} protein kinase encoded by *cdc2*⁺ and transcription factors encoded by *cdc10*⁺ and *res1*⁺/*scf1*⁺ (refs 7–10). Here we describe the characterization of the gene *rum1*⁺, which is important in G1 regulating Start. When *rum1*⁺ is overexpressed it breaks the dependence of Start upon completion of mitosis, initially leading to a transient G1 delay, followed by repeated rounds of DNA replication. When *rum1*⁺ is deleted, the pre-Start G1 interval is eliminated and the critical cell mass required for Start is reduced. This indicates that *rum1*⁺ is a major element determining the length of G1. *rum1*⁺ is also important for defining the cell in the pre-Start G1 interval: when *rum1*⁺ is deleted in a mutant that normally arrests before Start, cells proceed to undergo mitosis and cell division. These results indicate that the *rum1*⁺ gene product

is central for regulating G1 progression and the onset of S phase in fission yeast.

Overreplication

Certain mutant alleles of *cdc2*⁺ induce an extra round of DNA replication in G2 cells, breaking the normal dependence upon mitosis and generating cells with enlarged nuclei¹¹. This work led to the proposal that entry into S phase and mitosis is determined by the state of p34^{cdc2}. In order to identify new genes inducing overreplication, we used an *S. pombe* cDNA library (gift from B. Edgar and C. Norbury) under the control of the repressible *nmt1*⁺ promoter¹². Expression from this promoter begins at 12 h and becomes maximal at 14–16 h after transfer of cells to minimal medium lacking thiamine. The library was used to screen for clones that induced cells to generate enlarged nuclei. Among such clones we expected to find some that induced cells to undergo several rounds of DNA replication in the absence of mitosis. The most dramatic phenotype encountered of this type was associated with a 1.5-kilobase (kb) complementary DNA derived from a gene we called *rum1*⁺ (for replication uncoupled from mitosis). When *rum1*⁺ cDNA was overexpressed, it induced massive overreplication, producing highly enlarged nuclei (Fig. 1a, left). DNA content per cell estimated using flow cytometry was very much increased. An example of 16C cells is shown in Fig. 1a (right), but cells with still greater amounts of DNA were observed after prolonged exposure to elevated expression of *rum1*⁺.

The increase in DNA content was a result of complete rounds of DNA replication. During a time course of *rum1*⁺ overexpression, discrete subpopulations of 4C and 8C cells were observed (Fig. 1b), which would not be expected if there were only partial rounds of DNA replication. Also, if *rum1*⁺ was derepressed for

* Present address: Instituto de Microbiología-Bioquímica, CSIC/Universidad de Salamanca, Edificio Departamental, Campus Miguel de Unamuno, 37007 Salamanca, Spain.

only a short period, most surviving cells generated diploid or tetraploid clones, as would be expected if there were complete rounds of DNA replication. Therefore elevated *rum1*⁺ expression induces an extra S phase, breaking the normal dependence upon mitosis.

Although S phase dependence upon mitosis is disrupted, S phase is still coupled to cell-mass increase. Overreplicating haploid cells become highly enlarged (Fig. 1a, left), with both DNA and protein contents per cell increasing in parallel from 32.3 fg DNA per cell and 15.3 pg protein per cell at 0 h to 74.6 fg DNA per cell and 42.2 pg protein per cell by 22 h. The protein/DNA ratio was elevated by ~20% in the overreplicating cells, indicating that the cell mass at which S phase takes place is slightly increased. This increase in protein/DNA ratio was not simply due to the cells becoming polyploid, because the protein/DNA ratio was reduced by 13% in a diploid wild type compared with a haploid wild type. The increased protein/DNA ratio means that elevated *rum1*⁺ expression results in a small delay of S phase onset until cells attain a higher mass, suggesting the *rum1*⁺ gene product may act as a transient inhibitor of progression through G1 into S phase. The appearance of a small population of 1C cells at 16 h is consistent with this interpretation (Fig. 1b).

To test whether the increased DNA content was a consequence of extra rounds of DNA replication occurring in the complete absence of mitosis, we monitored overexpressing *rum1*⁺ cells for mitotic features. Cell number increase stopped by 16 h just before the elevated DNA content phenotype appeared. By this time, the proportion of the population with a condensed nucleus or a mitotic spindle typical of a mitotic cell had reduced to less than 1% (Fig. 1c). The proportion remained at this low level for the rest of the experiment, during which time the DNA content per cell more than doubled (Fig. 1c). A further feature of mitotic cells is an elevated p34^{cdc2}/p56^{cdc13} (cyclin B) protein kinase activity. This was monitored by assaying p56^{cdc13} immunoprecipitates for protein kinase activity. As shown in Fig. 1c, mitotic kinase activity dropped to a very low level during 16–22 h, about 11% of the value at 0 h before *rum1*⁺ was overexpressed, and 2% of the value in mitotically dividing cells (Fig. 1 legend).

Entry into pre-Start G1

These experiments establish that the cells undergoing extra rounds of DNA replication show no features of mitosis, as assessed cytologically or biochemically. This conclusion suggests that G2 cells are able to undergo DNA replication. Further support for this conclusion was obtained using the temperature-

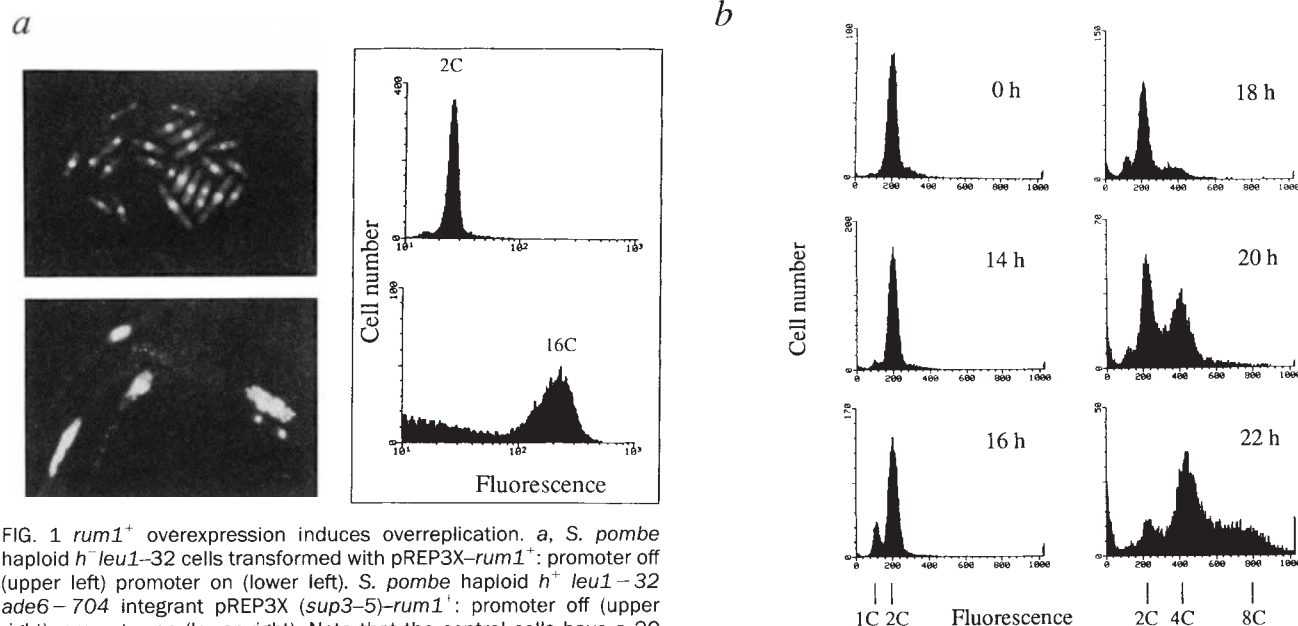
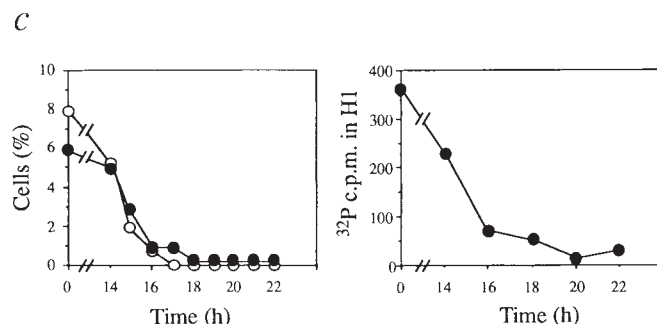


FIG. 1 *rum1*⁺ overexpression induces overreplication. *a*, *S. pombe* haploid *h*⁺ *leu1*-32 cells transformed with pREP3X-*rum1*⁺: promoter off (upper left) promoter on (lower left). *S. pombe* haploid *h*⁺ *leu1*-32 *ade6*-704 integrant pREP3X (*sup3*-5)-*rum1*⁺: promoter off (upper right), promoter on (lower right). Note that the control cells have a 2C DNA content. This is because in haploid wild-type fission yeast most of cell cycle is made up by G2, and both G1 and S phases are very rapid, with S being completed just as cells undergo cytokinesis. The G2 and M cells have a 2C DNA content, and the G1 cells have two 1C nuclei and thus also have a 2C DNA content¹. *b*, induction of *rum1*⁺ for 14 to 22 h in the integrant strain. FACS analysis with 1C, 2C, 4C, 8C peaks is indicated. *c*, Left, percentage of cells with mitotic condensed nuclei (●) and mitotic spindles (○), and right, mitotic p34^{cdc2}/p56^{cdc13} protein kinase activity using p56^{cdc13} immunoprecipitates as ³²P c.p.m. into histone H1; same time course as in *b*.

METHODS. An *h* *leu1*-32 strain transformed with pREP3X-*rum1*⁺ or an integrant constructed by transforming the strain *h*⁺ *ade6*-704 *leu1*-32 with the plasmid pREP3X(*sup3*-5)-*rum1*⁺, were grown to mid-exponential phase in minimal medium containing 5 μg ml⁻¹ thiamine at 32 °C (refs 37–39). Cells were washed 4 times with fresh minimal medium without thiamine, split into two cultures with 5 μg ml⁻¹ thiamine (promoter off) and without thiamine (promoter on), and grown at 32 °C. Cells were fixed in cold 70% ethanol, stained with DAPI (4',6-diamidino-2-phenylindole) and photographed using a Zeiss Axioscop photomicroscope. DNA content was determined by flow cytometry⁴⁰ (using a Becton-Dickinson FACScan) and by the diphenylamine method⁴¹ modified according to ref. 42. Protein content was determined using the BCA kit assay (Pierce). Cell number was measured using a Coulter counter. Nuclei were stained with DAPI, and mitotic spindles



with the anti-tubulin TAT1 antibody (gift from K. Gull) after fixation in formaldehyde and glutaraldehyde³⁷. Protein kinase was assayed after immunoprecipitation with anti-p56^{cdc13} (cyclin B) antibodies using histone H1 as a substrate⁴³. Radioactive ³²P incorporation into H1 histone was counted using a scintillation counter. As a control, activity was also assayed in mitotically dividing cells prepared by shifting a *cdc25*-22 culture to 36 °C for 4.5 h, followed by 20 min at 25 °C. Incorporation into H1 histone assayed as described was 1,970 c.p.m. (mean of 3 estimates).

sensitive mutant *cdc25-22* which blocks cells in G2 (refs 4, 13). *rum1⁺* was expressed in *cdc25-22* at 25 °C (Fig. 2b, right), the culture was split and half was shifted to the restrictive temperature of 36 °C for a further 1.5 generation times to arrest cells in G2. At both temperatures the DNA content per cell continued to increase, demonstrating that the G2-arrested cells were able to undergo further rounds of S phase (Fig. 2c and d, right).

The effect of *rum1⁺* overexpression could be to redirect G2 cells into G1, after which they would proceed to S phase. If this redirection occurred into the pre-Start G1 interval, then mutants that block progression through G1 at Start should also prevent the increase in DNA content. This possibility was tested using *cdc10^{ts}* and *cdc2^{ts}* mutants, which block at Start^{4,7}. *rum1⁺* was expressed in *cdc10-129* and *cdc2-L7* at the permissive temperature of 25 °C (Fig. 2b, left and middle). Cultures were then split and half was shifted to 36 °C for a further 1.5 generation times to arrest cells at Start. In both strains at 25 °C, DNA content continued to increase (Fig. 2d, left and middle), but at 36 °C where Start was blocked, there was almost no increase in DNA content (Fig. 2c, left and middle). This experiment indicates that elevated *rum1⁺* expression results in cells being redirected into G1 before Start (Fig. 2a).

Delaying G1 progression

These experiments show that increased *rum1⁺* expression disturbs normal G1 regulation. To address whether the *rum1⁺* gene product has a direct role in cell cycle regulation, we deleted the *rum1⁺* gene (Fig. 3 legend). Rapidly growing cells were

unaffected by the deletion of *rum1⁺*, but such cells have a very short G1 and lack a significant pre-Start G1 interval. In contrast, when wild-type cells are nitrogen-starved they accumulate in G1 before Start^{7,14} (Fig. 3a, top). When cells deleted for *rum1⁺* are nitrogen-starved, no G1 population is observed (Fig. 3a, bottom). Thus *rum1⁺* is required for the G1 arrest at Start that results from nitrogen starvation. Cells deleted for *rum1⁺* were also found to be sterile, possibly because they cannot accumulate before Start before conjugation.

A second situation in which the pre-Start G1 interval is extended is in the temperature-sensitive *wee1-50* mutant^{1,15}. Two hours after shifting to 36 °C, *wee1-50* cells begin to accumulate in G1 (Fig. 3a, top right) but no such accumulation was seen in *wee1-50* cells deleted for *rum1⁺* (Fig. 3a, bottom right). Both strains were grown at 25 °C and synchronous cultures were made using an elutriation rotor. After inoculation at 36 °C, both strains underwent cell division at the same time, as seen by the peaks in the percentage of dividing cells at 100 min in the two cultures (Fig. 3c). In the *wee1-50* culture, a G1 peak was observed, appearing at 100 min, rising to a maximum at 120 min when ~40% of the cells were in G1, and persisting until 160 min (Fig. 3b, top). In contrast, no G1 peak was seen at all in the *wee1-50* strain deleted for *rum1⁺* (Fig. 3b, bottom). Indeed, at 100 min about 30% of the cell population had a greater than 2C DNA content, indicating that the cells were initiating S phase even before cell division was completed. The double mutant dies after a few divisions, indicating that *rum1⁺* is essential in cells with an extended G1.

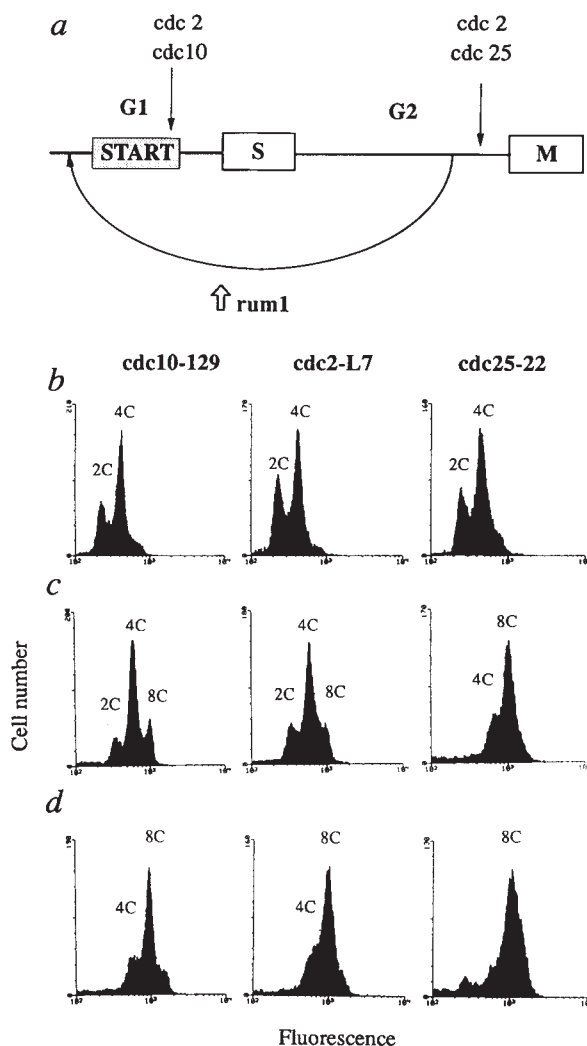


FIG. 2 Overexpression of *rum1⁺* redirects G2 cells to the pre-Start G1 interval. a, *cdc10-129* blocks at the restrictive temperature in G1 before Start, *cdc25-22* blocks in G2 before mitosis, and *cdc2-L7* blocks both in G1 before Start and G2 before mitosis. b, FACS profile of a strain containing an integrated copy of pREP3X(*sup3-5*)-*rum1⁺* in *cdc10-129*, *cdc2-L7* and *cdc25-22* mutant backgrounds after 24 h induction at 25 °C. c, FACS profile after 24-h induction at 20 °C followed by 4.5 h at 36 °C; and d, FACS profile after 30 h induction at 25 °C. As cells elongate, there is a gradual shift of the DNA peaks rightwards, probably as a consequence of increased staining in the enlarged cells due to mitochondrial DNA⁴⁰.

METHODS. An integrant was constructed by transforming the strain *h⁺ ade6-704 leu1-32* with the plasmid pREP3X(*sup3-5*)-*rum1⁺* (ref. 37). This integrant was crossed to the strains *cdc10-129*, *cdc2-L7* and *cdc25-22* to obtain double mutants. These double mutants were grown in minimal medium containing 5 µg ml⁻¹ thiamine at 25 °C to mid-exponential phase, washed 4 times with minimal medium and incubated in fresh minimal medium. After 24 h at 25 °C, half of the culture was shifted to 36 °C for 4.5 h (1.5 generations) and the other half was left at 25 °C for another 6 h (1.5 generations). Cells were fixed in cold 70% ethanol and DNA content was determined using a FACScan as for Fig. 1.

This experiment establishes that the G1 interval observed in a *wee1-50* strain is eliminated when *rum1*⁺ is deleted. Therefore *rum1*⁺ encodes an element determining the length of G1. In the *wee1-50* strain lacking the *rum1*⁺ gene product, the requirement to attain a critical cell mass before passing Start is slightly reduced. This was established by measuring the protein/DNA ratio, which was reduced by 18% when *rum1*⁺ was deleted. As a result, in the absence of *rum1*⁺ the *wee1-50* cells undergo Start and initiate S phase almost immediately after completing mitosis.

Defining pre-Start G1

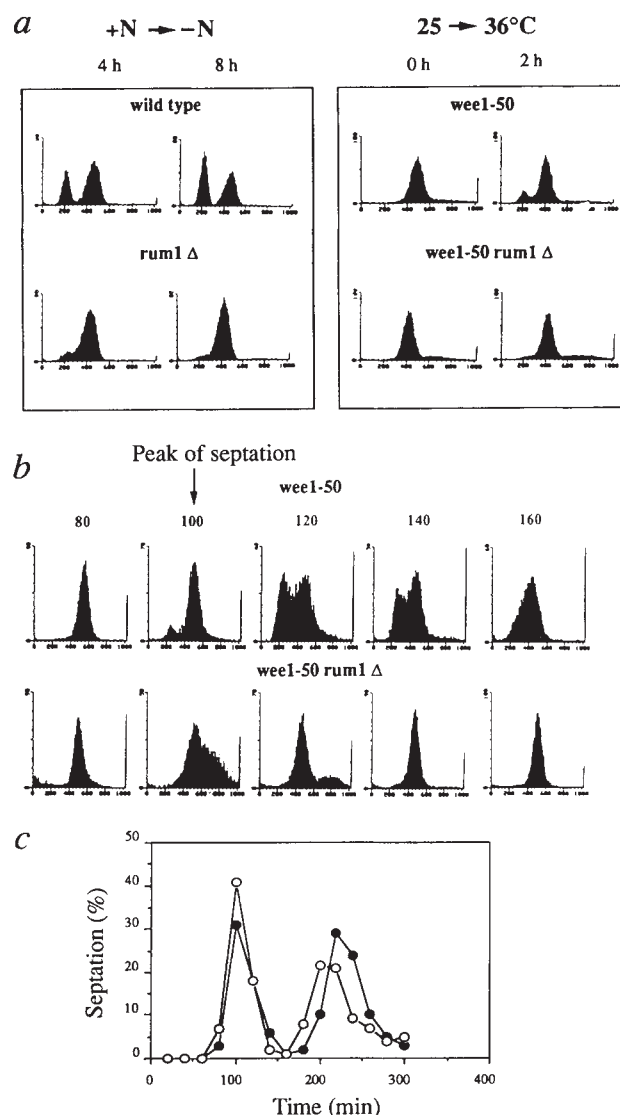
The fact that *rum1*⁺ overproduction causes redirection of a G2 cell into G1 before Start suggests that the *rum1*⁺ gene product has a role in defining this interval of the cell cycle. This was investigated by deleting *rum1*⁺ in the *cdc10-129* mutant. At 25 °C most cells in a *cdc10-129* culture are post-Start and upon shift to 36 °C proceed to divide once, resulting in a cell-number doubling before they arrest in the subsequent G1 (Fig. 4b, left). But *cdc10-129* cells deleted for *rum1*⁺ continue to divide further (Fig. 4b, left), generating a *cut*-like phenotype (aberrant mitoses and septation without mitosis)^{16,17} (Fig. 4b, right). This was examined further using a synchronous culture of the *cdc10-129* mutant deleted for *rum1*⁺. The synchronous culture was prepared at 25 °C and inoculated at 25 °C and 36 °C. At 25 °C the cells divided normally; two increases in the percentage of septation are shown in Fig. 4d (left). At 36 °C the first peak was

observed (Fig. 4f, left) because the cells used to inoculate the synchronous culture were in G2 and proceeded to divide once before arresting in G1 before Start with a 1C DNA content. If *rum1*⁺ was present, the *cdc10-129* cells would remain arrested in G1 for some hours and no more division would take place⁴ (data not shown). In the absence of *rum1*⁺, cells accumulated with a 1C DNA content but then proceeded to divide again, and by 8 h produced a population of cells with less than a 1C DNA content (Fig. 4f, right). This shows that *cdc10-129* cells deleted for *rum1*⁺ could not arrest in the pre-Start G1 interval and underwent mitosis and cell division even though S phase could not take place.

Premature entry into mitosis and cell division has been observed in certain *cdc2*⁺ mutants (e.g. *cdc2-3w* and *cdc2-F15*) blocked at the G1/S boundary with the DNA synthesis inhibitor hydroxyurea^{18,19}. The *cdc2-3w* mutant entered mitosis if cells were arrested in the post-Start G1 interval but not in the pre-Start G1 interval¹⁸, contrasting with the situation described here, where pre-Start-arrested cells enter mitosis. We tested whether cells deleted for *rum1*⁺ blocked post-Start with hydroxyurea would enter mitosis and cell division. Addition of hydroxyurea to wild-type and *rum1*⁺-deleted cells was identical in effect. Cell division stopped and cells became elongated with no mitotic features (Fig. 4c). When hydroxyurea was added to the G2 cells of the synchronous culture of the *cdc10-129* mutant deleted for *rum1*⁺, the first peak of dividing cells was seen (Fig. 4e, g). At

FIG. 3 *rum1*Δ mutants do not arrest in G1 in response to nitrogen starvation or in combination with a *wee1-50* mutant. a, FACS profile of the wild-type (upper) and *rum1*Δ (lower) strains 4 and 8 h after nitrogen starvation at 32 °C (left) and of the strains *wee1-50* (upper) and *wee1-50 rum1*Δ (lower) at 25 °C and 2 h after shift to 36 °C (right). b, FACS profile of synchronous cultures of *wee1-50* (upper) and *wee1-50 rum1*Δ (lower) at 36 °C. c, Septation index for the *wee1-50* (●) and the *wee1-50 rum1*Δ (○) synchronous cultures.

METHODS. A 2.2-kb *Clal* fragment in *rum1*⁺ was replaced with the *ura4*⁺ gene in a 6.1 kb *rum1*⁺ genomic clone (Fig. 5). This construction was used to delete the *rum1*⁺ gene using a diploid strain *h*⁺/*h*⁺ *ade6-M210/ade6-M216 leu1-32/leu1-32 ura4-d18/ura4-d18*. Stable *ura*⁺ diploids were obtained and tetrads were dissected after sporulation. Although there was a general reduction in the viability of the spores from these tetrads in comparison with a wild-type control, uracil prototrophic haploid colonies with the *rum1*⁺ gene deleted grew well in YE5S, indicating that the *rum1*⁺ gene was not essential in a wild-type background. For the nitrogen starvation experiment, cells were grown in minimal medium to mid-exponential phase, washed 6 times with nitrogen-free minimal medium and resuspended in this medium at a concentration of 2×10^6 cells per ml. The *rum1*Δ strain is sterile. To construct a double mutant with *wee1-50* it was necessary to transform the strain *h*⁺ *rum1::ura4*⁺ *ade6-M210 ura4-d18 leu1-32* with the plasmid pREP3X-*rum1*⁺ (promoter off) and perform the cross in the presence of 5 μg ml⁻¹ thiamine. The *wee1-50* and the *wee1-50 rum1*Δ strains were grown at 25 °C in minimal medium (where they were indistinguishable from wild-type cells) and then shifted to 36 °C. A synchronous culture of these two strains growing at 25 °C was prepared by centrifugal elutriation. The synchronized cells (in early G2) were then incubated at 36 °C, and the septation index and DNA content followed for two generations. Septation index was monitored in a Zeiss microscope using dark-field microscopy. DNA content and protein content was measured as indicated in Fig. 1.



25 °C the second peak of division was eliminated (Fig. 4e), but at 36 °C most of the cells went into mitosis with normal-looking spindles (data not shown), and then into cell division (Fig. 4g), generating cells with less than a 1C DNA content. This was because at 36 °C the cells reached the *cdc10-129* block-point first and so could proceed to mitosis and cell division as they lacked *rum1*⁺ function.

We conclude that the *rum1*⁺ gene product is important for defining a cell in the pre-Start G1 interval. When *rum1*⁺ is deleted, then a cell blocked in this interval cannot recognize the

situation and undergoes mitosis and cell division on schedule, even though the DNA is not replicated. If the same cells are allowed to proceed into the post-Start G1 interval, where progression is blocked using hydroxyurea the cell now recognizes the situation and prevents mitosis and cell division. These are examples of checkpoint controls by which a cell monitors whether earlier cell-cycle events have been completed before proceeding with later cell-cycle events^{20,21}. Our data indicate that two different checkpoint controls may operate in the G1 interval, one before Start involving *rum1*⁺, and a second post-Start which

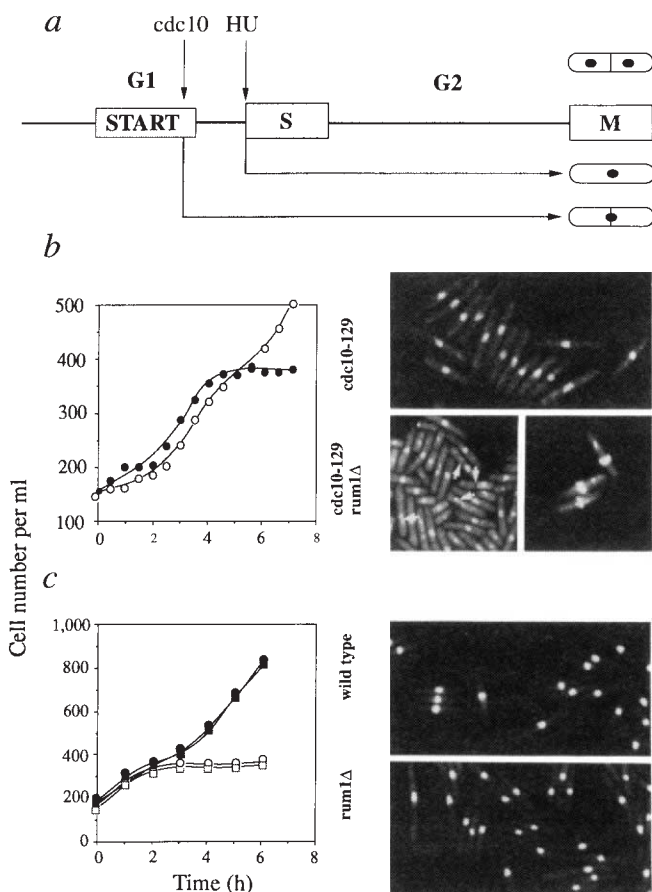
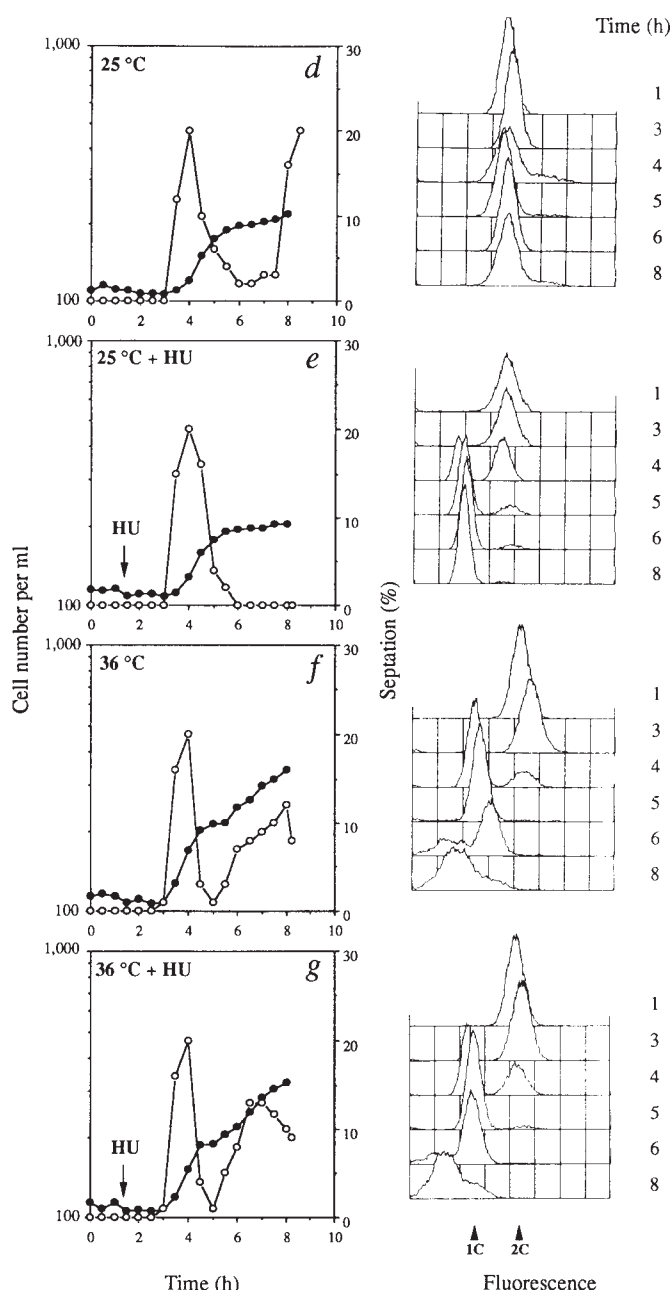


FIG. 4 *a*, *cdc10-129* mutant blocks at the restrictive temperature in G1 at Start. Hydroxyurea blocks the cell cycle post-Start at the initiation of S phase. *b*, cell number increase (left) and phenotype (right) of *cdc10-129* (●) and *cdc10-129 rum1Δ* (○) 5 h after the shift from 25 to 36 °C. Cells were stained with DAPI, arrows indicate cells that are undergoing aberrant mitosis or cell division. The *cdc10-129 rum1Δ* cells were also stained with DAPI and calcofluor to show the cut phenotype (lower right). *c*, Cell number increase (left) and phenotype (right) of wild-type (circles) and *rum1Δ* (squares) in the absence (filled symbols) and in the presence (open symbols) of 10 mM hydroxyurea (HU). DAPI-stained cells are shown in the presence of HU for 6 h. *d-g*, Cell number (●), septation index (○) (left) and DNA content (right) of synchronized *cdc10-129 rum1Δ* at 25 °C (*d*), 25 °C in the presence of hydroxyurea (*e*), 36 °C (*f*) and 36 °C in the presence of hydroxyurea (*g*).

METHODS. Wild-type and *rum1Δ* strains were grown in minimal medium at 32 °C. At a density of 2×10^6 cells per ml, the cultures were split into two, and half of the culture was treated with 10 mM hydroxyurea. Samples were taken every hour to monitor cell number increase using a Coulter counter and every 2 h to stain with DAPI. *cdc10-129* and *cdc10-129 rum1Δ::ura4⁺* were grown on minimal medium at 25 °C to mid-exponential phase. The cultures were shifted to 36 °C and cell-number increase was monitored every 30 min. Samples for FACS analysis and microscopic examination were taken every hour. A culture of the *cdc10-129 rum1Δ* strain was grown at 25 °C in minimal medium



and cells were synchronized using an elutriator rotor. These cells were split into four identical cultures, two were incubated at 25 °C and two at 36 °C. After 1 h and 45 min, hydroxyurea (or the equivalent volume of water) was added to a final concentration of 10 mM to one of the two cultures at each temperature. Samples were taken to measure cell-number increase, septation index and DNA content by FACS.

does not involve *rum1*⁺. If a wild-type cell is arrested in earlier of these two intervals, then the subsequent mitosis is blocked but by different mechanisms.

rum1⁺ gene function

We have determined the sequence of the *rum1*⁺ gene. Analysis of the nucleotide sequence identified an open reading frame with no introns which could potentially encode a protein of *M_r* 25K. This gene product is unlike any known proteins in the databases. It contains a putative bipartite nuclear localization sequence^{22,23}, suggesting that the protein may be targeted to the nucleus, and there are also consensus phosphorylation sites for both the Cdc2 and MAP protein kinases (Fig. 5)^{24,25}.

There are three aspects of *rum1*⁺ gene function involved in regulating the pre-Start G1 interval. First, it acts as a major element that determines the length of G1. It functions as a transient inhibitor of S-phase onset, delaying Start until the cell has attained a critical minimal mass. When *rum1*⁺ is deleted this mass is reduced, and when *rum1*⁺ is overexpressed this mass is increased. The only other class of genes that determine the length of G1 are the G1 cyclins. Dominant mutants in CLN genes shorten the G1 interval in budding yeast²⁶⁻²⁹. This is analogous to the effects of *rum1*⁺, except G1 cyclins act positively whereas *rum1*⁺ acts negatively. In mammalian cells ectopic expression of cyclin E can also shorten the G1 interval³⁰.

Second, *rum1*⁺ influences the dependence of Start and S phase upon completion of the previous mitosis. When *rum1*⁺ is overexpressed, G2 cells are redirected into the pre-Start G1 interval (Fig. 2), resulting in repeated rounds of DNA replication in the absence of mitosis. The checkpoint control that monitors whether mitosis has been completed is disturbed in these cells. Disturbance of the dependence of S phase upon completion of mitosis may occur naturally during endoreduplication in certain tissues in multicellular organisms such as *Drosophila*³¹. This dependence may also be influenced by licensing factor, which has been proposed as being necessary for S phase to take place in *Xenopus* eggs, but which can only gain entry to the nucleus during mitosis³². It has been hypothesized that licensing factor is consumed during the process of DNA replication, and so before another S phase can take place the cell must pass through mitosis. But the molecular nature of this factor remains unknown. The Chinese hamster *ts41* mutant also undergoes S phase without mitosis³².

Third, the *rum1*⁺ gene product defines the cell as being in the pre-Start G1 interval, and restrains such a cell from undergoing mitosis. When *rum1*⁺ is overexpressed, mitotic onset is inhibited, and when *rum1*⁺ is deleted, cells cannot arrest in the pre-Start G1 interval and undergo mitosis even though S phase has not taken place. The other gene involved in maintaining the dependence of S phase upon mitosis in fission yeast is *cdc2*⁺. When *cdc2* temperature-sensitive mutants are heat-shocked, G2 cells undergo S phase without mitosis¹¹. It was proposed on the basis of these experiments that a cell is defined in G2 or G1 by alternative forms of p34^{cdc2}. Our study here has also identified *rum1*⁺ as being important for defining a cell in the G1 interval, and preliminary studies have suggested that *cdc2*⁺ and *rum1*⁺ may interact in this control, because *rum1*⁺ seems to be required for the temperature-sensitive *cdc2* mutants to undergo S phase without mitosis (K. Labib, personal communication).

Pre-Start checkpoint

Our results indicate that in fission yeast there are two checkpoint mechanisms restraining mitosis, one operating pre-Start and a second post-Start. *rum1*⁺ restrains onset of mitosis when cells are in the pre-Start G1 interval. Fission yeast cells arrested post-Start send a signal that blocks p34^{cdc2} function at mitosis^{34,35}, this signal may be generated by the assembly of replication complexes post-Start, and the blocking of mitosis is influenced by the phosphorylation state of Tyr15 in the p34^{cdc2} protein kinase¹⁹. Cells arrested before Start may not have embarked

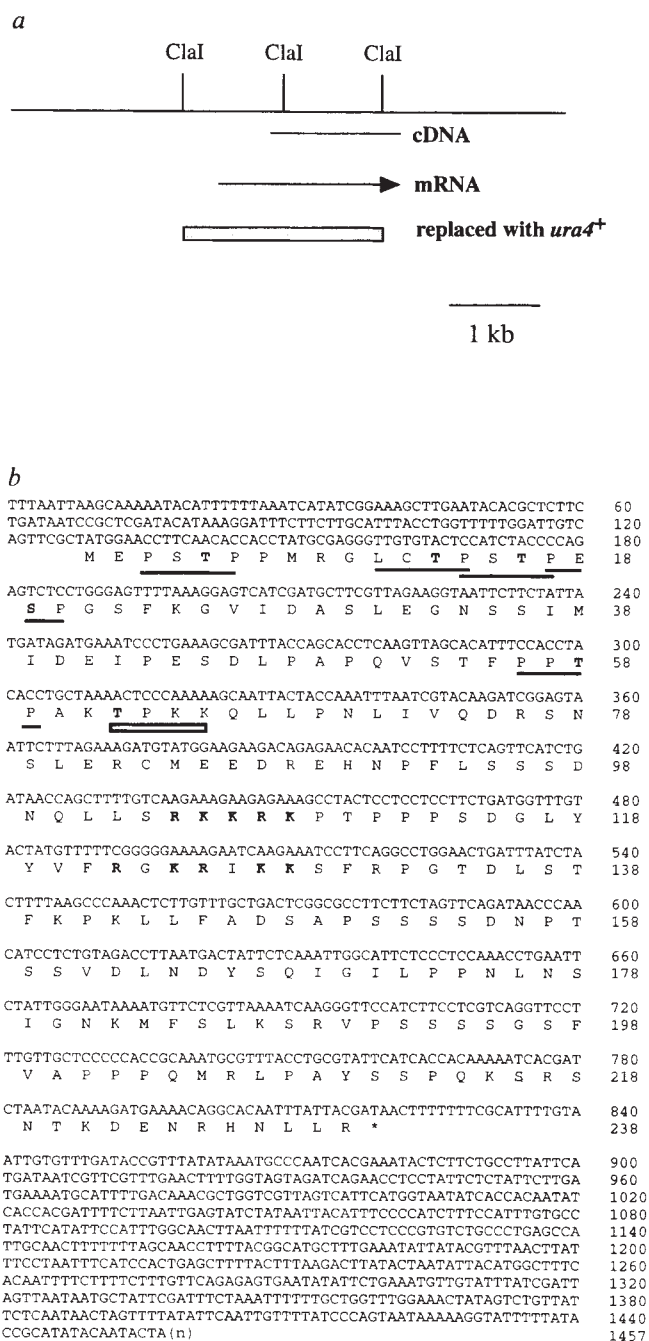
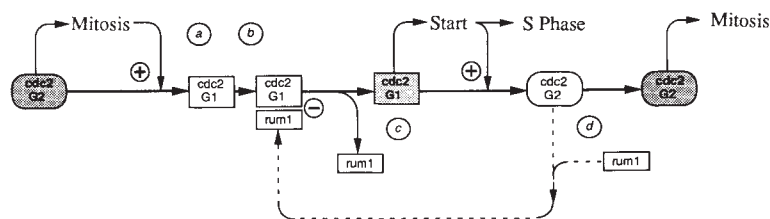


FIG. 5 a, Construction of the *rum1*⁺ deletion. The 2.2-kb *Cla*I restriction fragment in a 6.1-kb *rum1*⁺ genomic clone was replaced with the *ura4*⁺ gene and this construction was used to delete the *rum1*⁺ gene from the chromosome. b, DNA and protein sequence of the *rum1*⁺ gene. MAP kinase (bold underline) and Cdc2 kinase (bar underline) putative phosphorylation sites are indicated. A putative nuclear localization sequence is also indicated in heavy type. The *rum1*⁺ gene maps to chromosome II at 1 cM from the *cdc10* locus. The gene has been also mapped by hybridization to an ordered *S. pombe* cosmid library⁴⁴. The hybridization pattern is identical to that of the probe 4f12 (ref. 44). METHODS. The *rum1*⁺ cDNA was entirely sequenced on both strands using the dideoxy chain termination procedure of Sanger. Some parts of the sequence were confirmed using the equivalent genomic fragment. A 6.1-kb genomic fragment was obtained by hybridization of a genomic library using the *rum1*⁺ cDNA as a probe. The *rum1* gene was shown to be linked to *cdc10* by tetrads and random spore analysis.

FIG. 6 Passage through Start and entry into mitosis are brought about respectively by the G1/S and G2/M forms of $p34^{cdc2}$. The $rum1^+$ gene product maintains $p34^{cdc2}$ in the G1/S form and inhibits its action until the cell mass required for Start is attained. After Start, $p34^{cdc2}$ is converted to the G2/M form in preparation for mitosis. a, At completion of mitosis, $p34^{cdc2}$ is converted from the G2/M to G1/S form, possibly as a consequence of B-cyclin degradation. b, If the cell is of insufficient mass to undergo Start, then the $rum1^+$ gene product maintains $p34^{cdc2}$ in its G1/S form and inhibits its function until that minimal cell mass is attained. c, On attainment of the required minimal cell mass, the inhibitory effect of the $rum1^+$ gene product is lost, allowing Start to take place and $p34^{cdc2}$ to be converted to the G2/M form. In the absence of the $rum1^+$ gene product, Start and S phase occur prematurely at a reduced cell mass and the G2/M form of $p34^{cdc2}$ can be generated, resulting in premature mitosis. d, If there



are high levels of the $rum1^+$ gene product present, the G2/M form of $p34^{cdc2}$ cannot be formed and cells are redirected into the pre-Start G1 interval by the maintenance of $p34^{cdc2}$ in the G1/S form. Repeated cycles of this behaviour result in a block over mitosis and repeated rounds of S phase.

on assembly of replication complexes, and indeed the $cdc2-3w$ mutant that initiates mitosis when arrested post-Start does not initiate mitosis when arrested before Start¹⁸. Therefore there may be a separate checkpoint mechanism restraining mitosis in pre-Start cells and it is a role in such a control that we propose for $rum1^+$.

Model for $rum1^+$ action

Based on the above discussion and previous results concerning $p34^{cdc2}$ (ref. 11), we propose a tentative model for $rum1^+$ action illustrated in Fig. 6. On completion of mitosis, $p34^{cdc2}$ is converted from the G2/M to the G1/S form, possibly as a consequence of B-cyclin degradation at the end of mitosis. If the cell is already above the critical mass required for Start, then this step is completed immediately after mitosis and S phase is rapidly initiated. This is the situation in a rapidly growing wild-type cell. But if the cell is too small for Start, $p34^{cdc2}$ is maintained in the G1/S form by the $rum1^+$ gene product which inhibits $p34^{cdc2}$ Start function until the critical mass is attained. This stabilization of the G1/S form of $p34^{cdc2}$ prevents conversion into the

G2/M form and thus blocks mitotic onset. At the critical mass the $rum1^+$ inhibitory effect is lost and the cell executes Start and proceeds to S phase. At the same time $p34^{cdc2}$ is converted to the G2/M form but is restrained from bringing about mitosis by a new checkpoint control activated by the assembly of replication complexes working through $p34^{cdc2}$ Tyr 15 phosphorylation. Lack of $rum1^+$ produces a cell that is unable to delay Start, explaining why no G1 is detectable in cells deleted for $rum1^+$. In the absence of $rum1^+$, a cell cannot arrest pre-Start and the failure to inhibit $p34^{cdc2}$ leads to premature mitosis. If the $rum1^+$ gene product is present at high levels in a G2 cell, this results in the inappropriate stabilization of $p34^{cdc2}$ in the G1/S form, which redirects the cell into the pre-Start G1 interval.

This model is speculative but provides a conceptual framework for future experiments. But it is clear that the $rum1^+$ gene product influences three important aspects of G1 regulation, indicating that it plays a central role in regulating this part of the cell cycle in fission yeast. Given the conservation of cell cycle controls³⁶, similar gene functions may have related functions in other eukaryotic cells. □

Received 1 September; accepted 1 December 1993.

- Nurse, P. *Nature* **256**, 547–551 (1975).
- Nurse, P. & Thuriaux, P. *Expl Cell Res.* **107**, 365–375 (1977).
- Nasmyth, K., Nurse, P. & Fraser, R. J. *Cell Sci.* **39**, 215–233 (1979).
- Nurse, P., Thuriaux, P. & Nasmyth, K. *Molec. gen. Genet.* **146**, 167–178 (1976).
- Hartwell, L. *Bact. Rev.* **38**, 164–198 (1974).
- Hartwell, L. & Unger, M. J. *Cell. Biol.* **75**, 422–435 (1977).
- Nurse, P. & Bissett, Y. *Nature* **292**, 558–560 (1981).
- Lowndes, N., McInerney, L., Johnson, A., Fantes, P. & Johnston, L. *Nature* **355**, 449–452 (1992).
- Tanaka, K. et al. *EMBO J.* **11**, 4923–4932 (1992).
- Caligiuri, M. & Beach, D. *Cell* **72**, 607–619 (1993).
- Broek, D., Bartlett, R., Crawford, K. & Nurse, P. *Nature* **349**, 388–393 (1991).
- Maundrell, K. J. *biol. Chem.* **265**, 10857–10864 (1989).
- Thuriaux, P., Sipiczki, M. & Fantes, P. A. J. *gen. Microbiol.* **116**, 525–528 (1980).
- Egel, R., Nielsen, O. & Weiguny, D. *Trends Genet.* **6**, 369–373 (1990).
- Nurse, P. & Thuriaux, P. *Genetics* **96**, 627–637 (1980).
- Uemura, T. & Yanagida, M. *EMBO J.* **3**, 1737–1744 (1984).
- Hirano, T., Funahashi, S., Uemura, T. & Yanagida, M. *EMBO J.* **5**, 2973–2979 (1986).
- Enoch, T. & Nurse, P. *Cell* **60**, 665–673 (1990).
- Enoch, T., Carr, A. & Nurse, P. *Genes Dev.* **6**, 2035–2046 (1992).
- Hartwell, L. & Weinert, T. *Science* **248**, 629–634 (1989).
- Murray, A. *Nature* **359**, 599–604 (1992).
- Kalderon, D., Richardson, W. D., Markham, A. F. & Smith, A. E. *Nature* **311**, 33–38 (1984).
- Robbins, J., Dilworth, S. M., Laskey, R. A. & Dingwall, C. *Cell* **64**, 615–623 (1991).
- Moreno, S. & Nurse, P. *Cell* **61**, 549–551 (1990).

- Clarke-Lewis, I., Sanghera, J. S. & Pelech, S. L. J. *biol. Chem.* **266**, 15180–15184 (1991).
- Sudbery, P. E., Goodey, A. R. & Carter, B. L. A. *Nature* **288**, 401–404 (1980).
- Nash, R., Tokiwa, G., Anand, S., Erickson, K. & Futcher, A. B. *EMBO J.* **7**, 4335–4346 (1988).
- Cross, F. *Molec. cell. Biol.* **8**, 4675–4684 (1988).
- Hadwiger, J. A., Wittenberg, C., Richardson, H. E., De Barros Lopes, M. & Reed, S. I. *Proc. natn. Acad. Sci. U.S.A.* **86**, 6255–6259 (1989).
- Ohtsubo, M. & Roberts, J. M. *Science* **259**, 1908–1912 (1993).
- Smith, A. V. & Orr-Weaver, T. L. *Development* **112**, 997–1008 (1991).
- Blow, J. J. & Laskey, R. A. *Nature* **332**, 546–548 (1988).
- Handeli, S. & Weintraub, H. *Cell* **71**, 599–611 (1992).
- Kelly, T. J. et al. *Cell* **74**, 371–382 (1993).
- Saka, Y. & Yanagida, M. *Cell* **74**, 383–393 (1993).
- Nurse, P. *Nature* **344**, 503–508 (1990).
- Moreno, S., Klar, A. & Nurse, P. *Meth. Enzym.* **194**, 795–823 (1991).
- Hofer, F. et al. *Curr. Genet.* **1**, 45–61 (1979).
- Carr, A. M., MacNeill, S. A., Hayles, J. & Nurse, P. *Molec. gen. Genet.* **218**, 41–49 (1989).
- Sazer, S. & Sherwood, S. W. J. *Cell Sci.* **97**, 509–516 (1990).
- Bostock, F. *Expl Cell Res.* **60**, 16–26 (1970).
- Gendimenico, G. J., Bouquin, P. L. & Trampusch, K. M. *Analyt. Biochem.* **173**, 45–48 (1988).
- Moreno, S., Hayles, J. & Nurse, P. *Cell* **58**, 361–372 (1989).
- Hoheisel, J. D. et al. *Cell* **73**, 109–120 (1993).

ACKNOWLEDGEMENTS. We thank the cell cycle group for discussion. B. Edgar and C. Norbury for the *S. pombe* cDNA library, and A. Orfao, E. Becker, D. Martin-Zanca and P. Bolanos for their help. This project was supported by the ICRF, the MRC and the Royal Society.

The X-ray crystal structure of the membrane protein prostaglandin H₂ synthase-1

Daniel Picot, Patrick J. Loll & R. Michael Garavito*

Department of Biochemistry and Molecular Biology, The University of Chicago, 920 East 58th Street, Chicago, Illinois 60637, USA

The three-dimensional structure of prostaglandin H₂ synthase-1, an integral membrane protein, has been determined at 3.5 Å resolution by X-ray crystallography. This bifunctional enzyme comprises three independent folding units: an epidermal growth factor domain, a membrane-binding motif and an enzymatic domain. Two adjacent but spatially distinct active sites were found for its haem-dependent peroxidase and cyclooxygenase activities. The cyclooxygenase active site is created by a long, hydrophobic channel that is the site of non-steroidal anti-inflammatory drug binding. The conformation of the membrane-binding motif strongly suggests that the enzyme integrates into only one leaflet of the lipid bilayer and is thus a monotopic membrane protein.

EICOSANOID compounds such as prostaglandins, thromboxanes and leukotrienes are formed from the essential fatty acid, arachidonic acid, by the enzymes in the arachidonate cascade¹. The enzyme prostaglandin H₂ synthase (PGHS) catalyses the first committed step in the biosynthesis of prostaglandins, converting arachidonic acid to prostaglandin H₂ (PGH₂), which is then rapidly converted into one of several prostanoids. Acting locally in a paracrine or autocrine fashion, prostaglandins initiate and modulate cell and tissue responses involved in many physiological processes such as platelet aggregation, renin release and inflammation¹. Prostaglandin biosynthesis has also been implicated in the pathophysiology of cardiovascular disease, cancer and inflammatory diseases^{2–4}. Non-steroidal anti-inflammatory drugs (NSAIDs) such as aspirin, ibuprofen and indomethacin directly target PGHS^{5,6}. While the NSAID inhibition of PGHS activity is a useful tool in the treatment of inflammation and pyresis, recent studies on the prophylactic use of aspirin⁷ and fish oils⁸ to reduce the incidence of cardiovascular disease have generated additional interest in the modulation of prostaglandin biosynthesis, either through drug therapies or diet.

PGHS is an integral membrane protein located primarily in the endoplasmic reticulum⁶. PGHS is bifunctional: the initial cyclooxygenase reaction, the target for NSAIDs, converts arachidonic acid to PGG₂, while the subsequent peroxidase reaction converts PGG₂ to PGH₂. There are two isozymes, PGHS-1 and PGHS-2, which differ in their regulation of expression and tissue distribution^{9–11}. Ovine PGHS-1 is a 576-amino-acid, glycosylated protein with an apparent subunit relative molecular mass of 70,000 (refs 6, 12). PGHS-1 is considered to be involved in cell-cell signalling and maintaining tissue homeostasis¹ whereas PGHS-2 expression occurs in a limited number of cell types and is regulated by specific stimulatory events^{10,13,14}, leading to the hypothesis that PGHS-2 is responsible for the prostanoid biosynthesis involved in inflammation and mitogenesis¹¹. The discovery of the PGHS-2 isoform has opened up the possibility of the design of isozyme-specific drugs. We present here the 3.5 Å resolution X-ray structure of PGHS-1 complexed with the NSAID flurbiprofen, using crystals grown in the presence of nonionic detergents^{15,16} (Table 1).

Structure determination

Data sets were collected from native and derivative crystals using a laboratory source, although data beyond 3.4 Å exhibited poor

signal-to-noise ratios. Strong diffraction data to 2.8 Å resolution have been observed at beamline X8C at the National Synchrotron Light Source (Brookhaven National Laboratory); however, the crystals decay rapidly in the intense X-ray beam, necessitating further work before high-resolution data sets can be measured. The MIR phasing (Table 1), molecular interpretation (Fig. 1a) and preliminary refinement were therefore done at 3.5 Å resolution.

General description of the structure

The asymmetric unit contains two identical monomers related by a non-crystallographic two-fold symmetry axis (Fig. 1b). The two monomers exhibit extensive contacts, in agreement with the proposed dimeric structure of the detergent-solubilized protein⁶. The dimer is ellipsoidal with dimensions of 99 × 65 × 55 Å. Mature ovine PGHS-1 extends over residues 25–600 (residues 1–24 form the cleaved signal sequence), and interpretable electron density is seen for residues 33–586. The polypeptide chain has a high helical content (about 40% of the residues; Table 2) and forms almost no organized β-pleated sheet. The tertiary structure can be subdivided into three distinct folding units (Fig. 1c). The amino-terminal residues 34–72 form a small compact domain, held together by three intra-domain disulphide bonds. The overall conformation of this domain, including the arrangement of the disulphide bonds, is very similar to that of the epidermal growth factor (EGF)¹⁷. This EGF-like module is covalently linked to the main body of the enzyme by another disulphide bridge (Cys 37–Cys 159).

The second folding unit (residues 73–116) consists of a right-handed spiral of α-helices A, B, C and D along one side of the monomer (Fig. 1c). These helices are highly amphipathic and create a novel motif for insertion of the enzyme into the lipid bilayer (see below). The transitional helix, D, leads into the main body of the enzyme (Fig. 1c), which forms the third folding unit: a large, globular catalytic domain (residues 117–587).

The catalytic domain is a globular structure containing the cyclooxygenase and peroxidase active sites. It comprises two distinct, albeit intertwined, lobes of polypeptide chain (Fig. 1d). The larger lobe is built up around a 'V'-shaped structure formed by helices H5 and H6 which grips helix H2 like tweezers. Helices H3, H10, H18, H19 and intervening loops pack around this core of helices to complete the large lobe. The smaller lobe of the catalytic domain is created by the helices H1, H8, H12, H14, H15 and H16 (Fig. 1d). These six helices form a bundle with their axes roughly parallel. The two lobes are connected by six polypeptide segments, including helix H17.

* To whom correspondence should be addressed.

The peroxidase active site

The peroxidase activity in PGHS has two functions: it reduces the PGG₂ produced by the cyclooxygenase step and activates the cyclooxygenase reaction. The PGHS peroxidase activity appears to be a typical haem-dependent peroxidase, reducing the hydroperoxide substrate by a two-electron transfer reaction that gives rise to an oxo-ferryl porphyrin cation radical known as compound I^{18,19}. The oxidized haem returns to its resting state by two sequential one-electron transfers to reducing substrates, or by oxygen transfer to aryl sulphide compounds¹⁸. PGHS preferentially reduces PGG₂ and other long-chain fatty acids and aliphatic hydroperoxides.

TABLE 1 Data collection and phasing statistics

	No. of crystals	No. of measurements	No. of reflections	% Unique	Resolution Å	<i>R</i> _{merge} %
Native	4	112,745	32,349	96.3	3.4	7.6
AuCN	2	76,324	31,031	92.1	3.4	8.9
EMTS	8	76,809	24,056	59.0	3.4	8.1
PTCL	4	63,604	27,286	66.9	3.4	7.9

Heavy atom	No. of sites	<i>R</i> derivative†	<i>R</i> _{cutis} ‡	Phasing power§
		%		
AuCN	4	16.1	0.51	1.7
EMTS	4	19.1	0.63	1.3
PTCL	2	13.5	0.79	0.8

Resolution (Å)	10.7	8.2	6.7	5.7	4.9	4.3	3.9	3.5	Total
Number	154	214	273	324	379	428	459	463	2,684
centric									
FOM centric	0.77	0.83	0.84	0.77	0.73	0.67	0.64	0.50	0.69
Number	590	1,099	1,784	2,600	3,605	4,707	5,965	6,927	27,277
acentric									
FOM acentric	0.64	0.68	0.71	0.70	0.63	0.54	0.48	0.39	0.54

Crystal preparation. PGHS-1 from sheep seminal vesicles was purified and crystallized as described elsewhere (refs 15, 16 and D.P., P.J.L., J.E. Harlan and R.M.G., manuscript in preparation). A 10 mg ml⁻¹ solution of freshly reconstituted holoenzyme was dialysed overnight versus 20 mM sodium phosphate pH 6.7, 100–200 mM NaCl, 0.6% (w/v) β-octyl glucopyranoside and 0.1 mM flurbiprofen. The protein solution was then made 2–4% in polyethylene glycol (PEG) 4000 and set up in hanging-drop vapour diffusion trays against 2.4× concentrated buffer and 4–8% PEG 4000. Long, brown rod-shaped crystals (1.5 mm × 0.25 mm × 0.25 mm) grew within 1–2 weeks. The crystals are difficult to manipulate and easily disordered during soaking and mounting. **Data collection.** Preliminary analysis revealed the space group of the crystals to be either I222 or I2₁2₁2₁ (*a* = 99.4 Å, *b* = 210.3 Å, *c* = 233.1 Å). Data were collected with a FAST area detector mounted on an Elliott GX-21 rotating anode with a 300 μm focus cup and running at 40 kV and 70 mA. Typical crystal data collection parameters include 85 mm crystal-to-detector distance and 240 s exposure time per 0.1 degree frame. The crystals were cooled to 4 °C with a chilled air stream to increase the crystal lifetime. For each separate scan, reflections were indexed and integrated using the programs MADNES⁴⁸ and PROCOR⁴⁹. The individual scans were then merged and further processed using the CCP4 program suite. **Phasing.** Patterson maps using (*F*_{HA} – *F*_{NAT})² were used to determine the positions of the major heavy atom (HA) sites. Initial positions for three HA derivatives, sodium ethylmercurithiosalicylate (EMTS), K₂PtCl₆ and Au(CN)₃ were determined and the identity of the space group established as I222. Cross-difference and phase residual Fourier maps were used to search for minor peaks and to improve initial estimates of HA positions. A rotation function analysis and analysis of the distributions of HA sites independently revealed a noncrystallographic two-fold axis oriented within 5° of the 011 crystal axis. Heavy-atom parameters were refined using MLPHARE from CCP4 version 2.1.4 (mean figure of merit 0.54) and electron density of 3.5 Å resolution was calculated using figure of merit weighting. Phase improvement was achieved by solvent flattening in SQUASH³⁹ (final mean figure of merit 0.78), estimating solvent content (including the detergent regions) at 72%. An atomic model of PGHS-1 was built from electron density maps that were solvent flattened or solvent flattened followed by two-fold molecular averaging.

* $R_{\text{merge}} = \sum_i |I_i - \langle I \rangle| / \sum_i I_i$, where *I*_i is the intensity of an individual reflection and $\langle I \rangle$ the mean intensity of that reflection.

† $R_{\text{derivative}} = \Delta F / F = \sum_i |F_{\text{PH}} - F_P| / \sum_i |F_P|$ is the relative isomorphous difference between the native and derivative structure factor amplitudes (*F*_P and *F*_{PH}, respectively).

‡ $R_{\text{cutis}} = \sum_i |F_{\text{PH}} \pm F_P| - F_H| / \sum_i |F_{\text{PH}} - F_P|$, where *F*_H is the calculated heavy atom structure factor contribution; the summation is done for the centric reflections used in the heavy atom parameters refinement cycle.

§ Phasing power = $\langle F_H \rangle / \langle E \rangle$ where $\langle F_H \rangle$ is the r.m.s. of the heavy atom scattering factor and $\langle E \rangle$ is the r.m.s. lack of closure; the summation is done for the reflections used in the heavy atom parameters refinement cycle.

The peroxidase active site (Fig. 1*d*) is found at the interface between the large and small lobes of the catalytic domain, in a shallow cleft containing the high-spin Fe(III)-protoporphyrin IX prosthetic group. The proximal face of the haem packs against part of the small lobe; specifically, helices H12 and H8 and the residues following helix H8, including the proximal iron ligand His 388 (Fig. 2*a*). On the distal side of the haem, Gln 203 and the distal histidine, His 207, lie against the haem face, approaching to within ~5 Å of the haem iron (Fig. 1*a*), but they do not coordinate the metal. Residues 289–295, located before helix H5, form a small shield above His 207 and Gln 203, thus partially enclosing the peroxidase active site (Fig. 2*a*). Nevertheless, the shallow cleft exposes a large part of the haem to solvent, possibly to accommodate relatively large hydroperoxide substrates such as PGG₂. A loop of residues (211–220) following helix H2 protrudes from the surface of the protein and creates two narrow grooves which may be binding sites for PGG₂ and the reducing substrate. One groove extends outward from the distal side of the haem and the other extends along helix H12, directed away from the edge of the haem between the two propionic groups.

Homology to other peroxidases

The X-ray crystal structures of three other haem-containing peroxidases have been determined: yeast cytochrome *c* peroxidase (CCP)²⁰, a fungal lignin peroxidase (LiP)²¹ and canine myeloperoxidase (MPO)²². Comparison of PGHS with these peroxidases to assess the roles of different active site residues is therefore possible.

The sequence identity between any two of the known mammalian peroxidase sequences (MPO, eosinophil peroxidase, lactoperoxidase and thyroid peroxidase) is about 70%²², while the level of sequence identity between PGHS and any member of this set is only around 20%²². Nonetheless, the overall topology of PGHS in the catalytic domain is strikingly similar to that of MPO (Fig. 2*b* and *c*) and the two structures show greater similarity than would be predicted from their sequence homology. The major helices involved in haem binding in both enzymes (H2, H5, H6, H8 and H12) can be superimposed with an r.m.s. difference in Cα positions of 1.4 Å. The structural homology extends over the entire catalytic domain of PGHS; however, the regions falling outside the catalytic domain, that is, the membrane-binding motif and the EGF domain, have no equivalents in MPO.

PGHS is not structurally homologous to MPO alone, but to CCP and LiP as well. CCP and LiP share 18% sequence identity, but have very homologous structures²¹. They have no obvious sequence homology with PGHS or MPO, but helices H2, H5, H6, H8, H12, H16 and H17 have structurally equivalent helices in CCP and LiP (Fig. 2*b–d*). These helices have the same topological and spatial organization, display similar packing arrangements, and form the core of the catalytic domains in all four peroxidases.

The haem orientation in the binding pocket differs between the mammalian and non-mammalian peroxidases (Fig. 2*b–d*): the propionic groups point towards the amino-terminus of helix B in CCP and LiP²¹, but toward the carboxy-terminus of the equivalent helix, H2, in MPO²² and PGHS. Despite this difference, the proximal histidine is always located on or just after the equivalent of helix H8 (helix F in CCP and LiP), while the distal histidine is always located on the equivalent of helix H2. Arg 47 of CCP and Arg 43 of LiP, found on helix B one turn before the distal histidine, have been implicated in the catalytic mechanism. The spatially equivalent positions on helix H2 in PGHS and MPO are occupied by Gln 203 and Gln 91, respectively. The exact function of the active-site arginine in CCP is unclear, as mutation of Arg 47 to lysine or leucine has only limited effects on the enzyme's catalytic capability²³. Although Arg 239 in MPO lies roughly 4 Å away from the distal histidine²², there is no spatially equivalent arginine in PGHS or CCP. No active-site arginine is found on the distal side of the

haem in catalase, even though, like the peroxidases, this enzyme creates a compound I intermediate during catalysis. In this case, the catalytic role of arginine is filled by Asn 147 (ref. 24). It is, however, too early to conclude if Gln 203 of PGHS serves an equivalent catalytic role.

The cyclooxygenase active site

The cyclooxygenase active site consists of a long, narrow channel ($\sim 8 \times 25 \text{ \AA}$) extending from the outer surface of the membrane-binding motif through helices A, B and C to the centre of the PGHS monomer (Fig. 3a). Tyr 385 is found at the apex of the

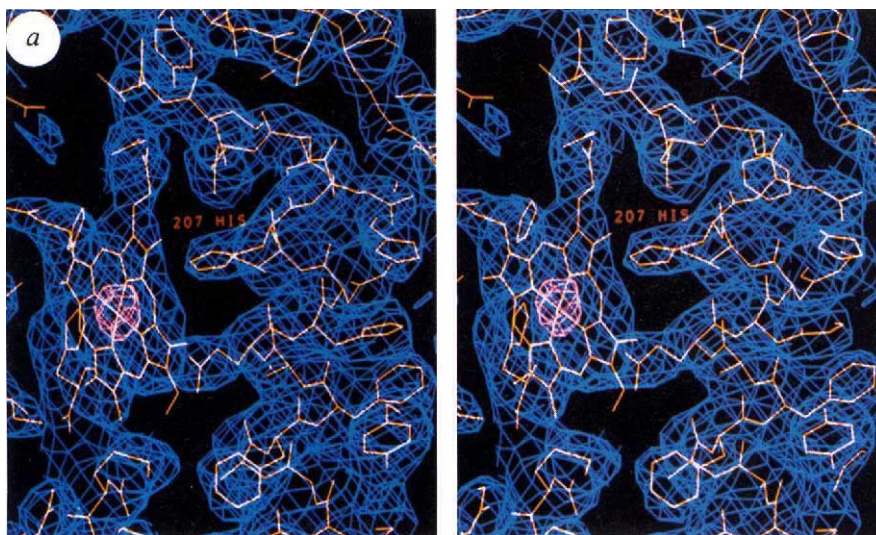
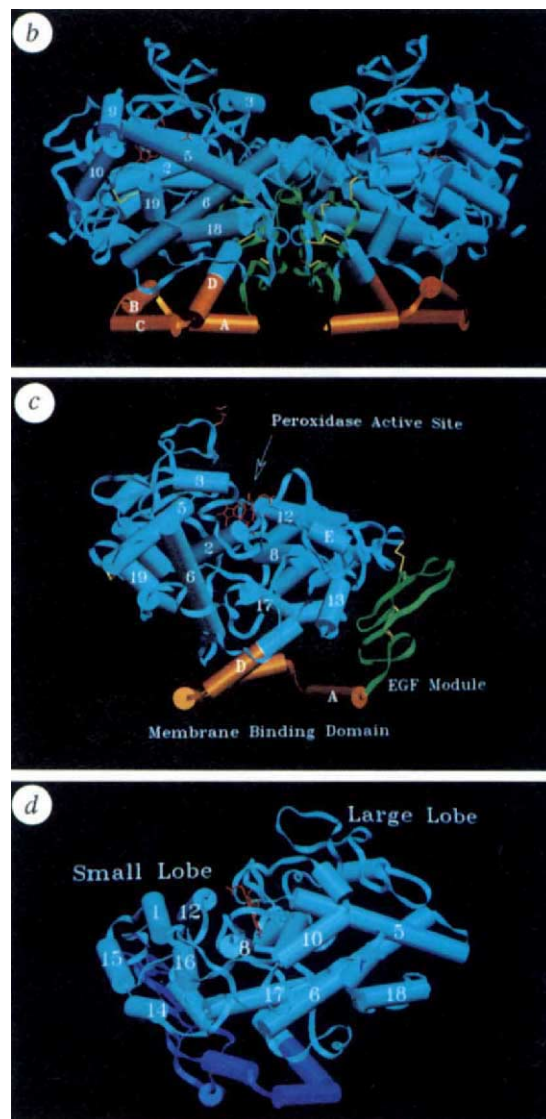


FIG. 1 a, Stereo view of the 3.5-Å electron density map in the vicinity of the haem pocket. The electron density was calculated with MIR phases improved by solvent flattening; the resulting electron density was then averaged about the non-crystallographic molecular two-fold axis and contoured at 1.1 σ . An atomic model of the enzyme was built into the electron density maps using FRODO⁴⁰ on an Evans and Sutherland PS390. The Luthy-Bowie-Eisenberg program gave scores typical for a protein of this size⁴¹. The initial atomic model was refined with XPLOR⁴² using data at 15–3.5 Å resolution. The initial and free *R*-values were both 45%; after simulated annealing these *R*-values were 26.7% and 31.6%, respectively. This partially refined model has r.m.s. deviations from ideality of 0.014 Å for bond lengths and 3.2° for bond angles. The haem-binding pocket is viewed from the solvent; note that the edge of the haem is at the surface of the protein. The distal histidine 207 extends out from the helix H2 towards the haem plane. One of the haem propionic groups faces towards the end of helix H2, while the other faces toward the opening of the active site. A two-fold ambiguity remains in the haem orientation, owing to an inability to distinguish between methyl and vinyl groups in the map. An anomalous difference Fourier map, calculated using the native data and MIR phases, clearly confirms the haem iron position (the 6 σ contour is shown in pink). b–d, Ribbon representation of the PGHS-1 dimer and monomer showing all the helices in the model; helices have been labelled according to Table 2. b, A dimer viewed along the dimer interface with the molecular two-fold axis vertical. The EGF-module, membrane-binding motif and catalytic domain are coloured green, rust and light blue, respectively; the haem is red and disulphide bonds (Cys 36–Cys 47, Cys 37–Cys 159, Cys 41–Cys 57, Cys 59–Cys 69 and Cys 569–Cys 575) are yellow. A monomer's contribution to the dimer interface includes an EGF-like module, helix E and the loop connecting helices D and E; the complementary surface provided by its symmetry mate includes parts of helices H5 and H6 and the intervening loop, a stretch of residues between H6 and H8, and the loop connecting helices H17 and H18. c, The left-hand monomer of a is rotated 90° about a vertical axis and viewed from the dimer interface to show the domain structure of the enzyme, using the same colour scheme as for b. The general structural features are labelled. The trypsin cleavage site^{16,43}, Arg 277, is on a loop above the peroxidase active site. His 309, once suggested to be a haem ligand⁴⁴, is located near the middle of helix H5, far from the haem. d, The monomer is rotated almost 180° about a vertical axis relative to c. In this orientation, the small and large lobes of the catalytic domain which create the peroxidase active site cleft are clearly visible. The EGF-like module, the membrane-binding domain and disulphide bridges are coloured dark blue to highlight the catalytic domain (in light blue). Figures 1b–d, 2, 3b and 4a–e were created with SETOR⁴⁵.



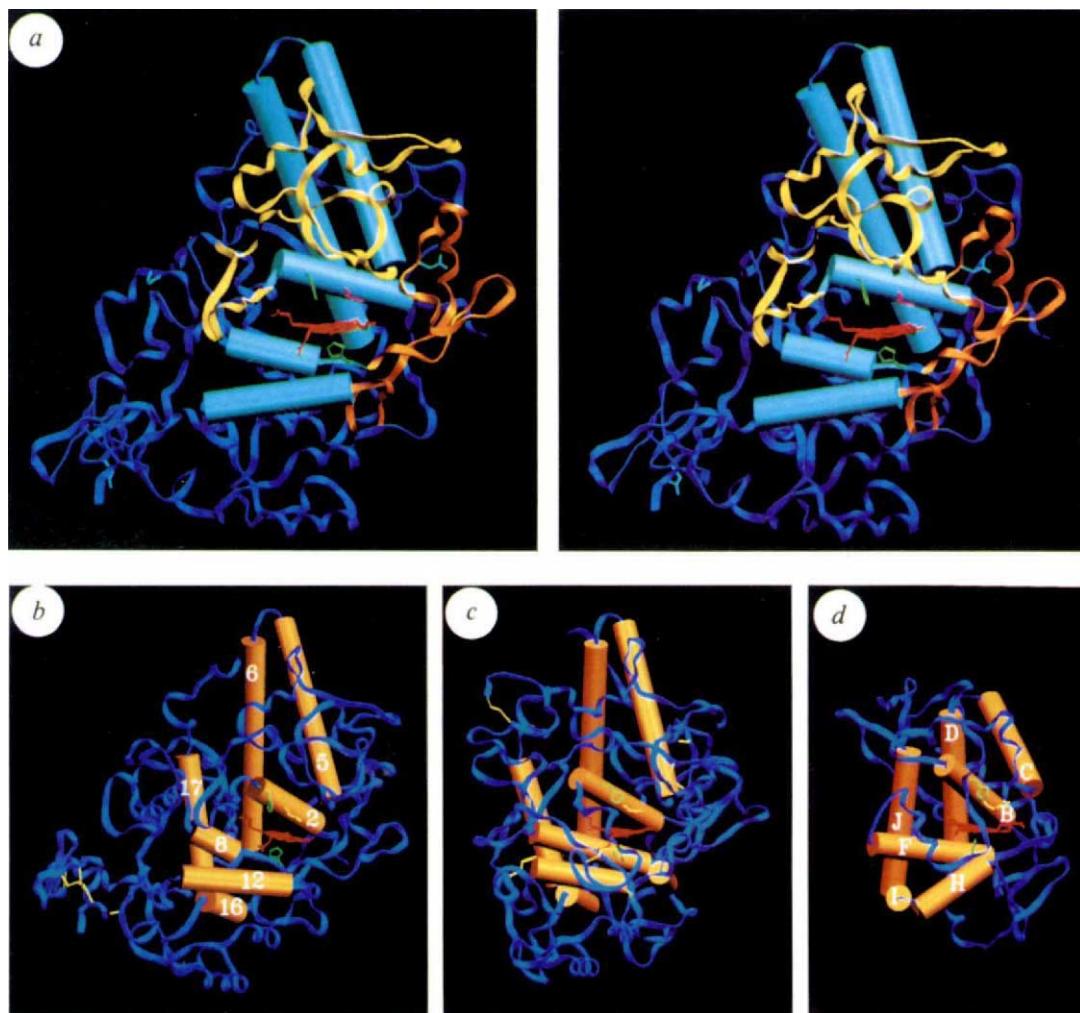
channel at the end of helix H8; it sits near the edge of the haem plane with its C γ within 10 Å of the haem iron and is surrounded by an annulus of six aromatic residues (not shown). Ser 530, which is known to be acetylated by aspirin⁶, lies just below Tyr 385 at a point where its acetylation could easily block access to the upper part of the channel (Fig. 3*b*). The channel also contains electron density that we have assigned to flurbiprofen, an NSAID added to the enzyme before crystallization. PGHS binds *S*-flurbiprofen stereospecifically⁶ and docking *S*-flurbiprofen into the density was straightforward. The position and orientation of flurbiprofen are confirmed by a low-resolution structure of PGHS complexed with an iodinated compound isosteric with flurbiprofen (P.J.L., D.P. and R.M.G., manuscript in preparation). The carboxyl group of the flurbiprofen is directed towards the mouth of the channel and lies in a favourable position for interacting with the guanidinium group of Arg 120 (Fig. 3*b*). Glu 524 is located near Arg 120, and these two residues may form a salt bridge; Arg 120 and Glu 524 are the only polar residues in the otherwise hydrophobic channel.

The cyclooxygenase active centre consists of the upper portion of the cyclooxygenase channel near Tyr 385 as deduced from the locations of Ser 530 and flurbiprofen. The existence of Arg 120 in the channel also provides a suitable moiety for providing a ligand to the carboxylate group of the substrate. It is not possible to construct a structure-based model of the reaction mechanism

from our limited resolution structure of a single enzyme-inhibitor complex; however, it is possible to build a stereochemically reasonable model of arachidonic acid into the active centre. The arachidonate would lie in a bent conformation with C-13 in the vicinity of Tyr 385 and the carboxylate group interacting with Arg 120. The architecture of the active site is then consistent with the proposed cyclooxygenase mechanism^{18,25,26}: PGHS first produces a radical on C-13 of arachidonic acid by the stereospecific removal of the pro-*S* hydrogen atom, allowing the subsequent insertion of molecular oxygen at C-11. The resulting 11-peroxyl radical initiates cyclization events that produce the 9,11-endoperoxide into which another molecule of oxygen inserts to form the 15-peroxyl radical. The donation of the hydrogen atom back to the substrate creates PGG₂. A tyrosyl radical has been detected leading to the proposal that the peroxidase activates the cyclooxygenase by radical transfer from the ferryl-oxo porphyrin cation radical of intermediate I (that is, compound I) to a tyrosine residue 9–12 Å away which then removes a hydrogen atom from arachidonic acid^{27,28}. The mutation of Tyr 385 to a phenylalanine abolishes the cyclooxygenase activity²⁹; its position in the cyclooxygenase site, its orientation relative to the haem and the putative arachidonic acid binding site are all consistent with this hypothesis, although its exact role in the catalytic process is unknown²⁶.

The current model of enzyme action also suggests a general

FIG. 2 A stereo ribbon representation of the PGHS-1 monomer highlighting the peroxidase active site is shown in *a*. The viewing direction is parallel to the molecular two-fold axis. The peroxidase active site is defined by the core helices H2, H5, H6, H8, H12 (coloured light blue), the H2–H5 intervening segment (yellow) and the H8–H12 intervening segment (rust). His 388 (the proximal iron ligand) and His 207 (the distal histidine) are shown in green. The haem is red and Gln 203 purple. Electron density is observed for at least the first saccharide unit at each of the three sites of *N*-linked glycosylation; these three sites are shown in light blue (Asn 68, Asn 144 and Asn 410)¹². *b*, PGHS-1; *c*, MPO; and *d*, CCP are shown in equivalent orientations to highlight their structural homology. Homologous helices are shown in rust (H2, H5, H6, H8, H12, H16, H17 in PGHS-1 and MPO; B, C, D, F, H, I, J in CCP); the proximal ligand (His 336 in MPO; His 175 in CCP) and distal histidines (His 95 in MPO; His 52 in CCP) are in green. Residues Gln 203 in PGHS-1, Gln 91 in MPO and Arg 48 in CCP are yellow, as are the disulphide bridges of PGHS and MPO. Small differences in the organization of secondary structural elements in MPO prevent the formation of a structural equivalent to the cyclooxygenase channel in PGHS. The haem in MPO



is also more buried than in PGHS; this is achieved by the locally folded region LR1²², which is located at the amino terminus of the large polypeptide in MPO and has no equivalent in PGHS.

explanation for NSAID inhibition. Whereas many NSAIDs are competitive inhibitors of the PGHS, flurbiprofen is part of a class of NSAIDs that induce a time-dependent inhibition of the cyclooxygenase⁶. It is not clear if this time-dependent phenomenon is a consequence of binding within a long, narrow channel or whether the inhibitor induces an irreversible change of the enzyme structure. Nevertheless, it can be argued that flurbiprofen inhibits the cyclooxygenase reaction by excluding arachidonate from the upper portion of the channel. This argument may explain the action of most, if not all, NSAIDs, but the extreme diversity of NSAID structures³⁰ and the narrow shape of the channel suggest that a number of subsites, some kinetically indistinguishable, could exist for drug binding.

Interaction with the lipid bilayer

PGHS is classified as an integral membrane protein because detergents, and not merely chaotropic salts, are required to extract the enzyme from the membrane⁶. Furthermore, proteo-

lytic cleavage has failed to yield an active soluble fragment, suggesting that the enzyme is not simply tethered to the membrane by a flexible anchor. Since the determination of the PGHS-1 amino-acid sequence^{31,32}, several models have been proposed requiring one or more transmembrane segments^{13,31,32}. These models are incompatible with the X-ray structure, as the proposed transmembrane segments are part of the catalytic domain. It is highly unlikely that detergent extraction or crystallization induces a major conformational change which completely alters the native, membrane-bound structure, because the catalytic domain is structurally homologous with the soluble enzyme MPO.

The X-ray structure offers an attractive model to explain the enzyme's interaction with a membrane. The enzyme has a compact structure with the catalytic domain clearly similar to that of MPO. The surface of this domain is polar and should therefore be external to the membrane. The three sites of *N*-linked glycosylation and the five disulphide bridges should be found

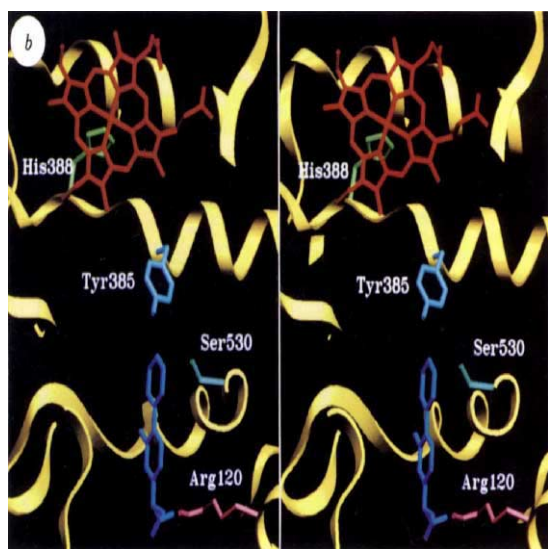
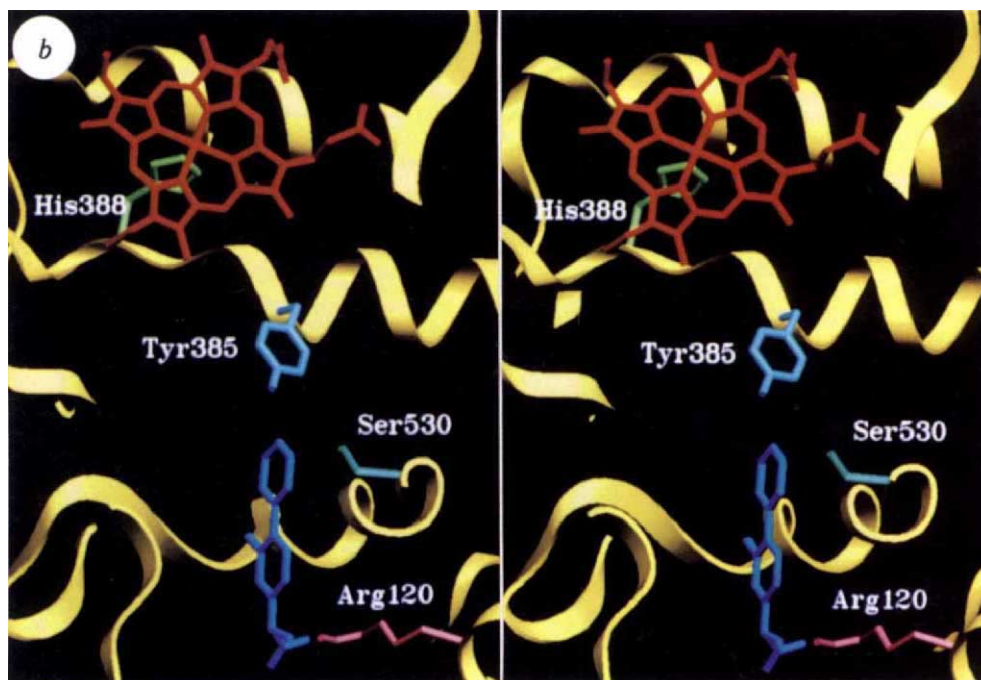


FIG. 3 Views of the cyclooxygenase channel and active site. *a*, A van der Waals surface shows the hydrophobic cyclooxygenase channel extending within a monomer (in blue) from the membrane-binding motif at the bottom to Tyr 385 (in greenish yellow) at the top, just underneath the haem (in red). The long axis of the channel passes through the intersection of helices D, H6, H8 and H17 between the large and small lobes of the catalytic domain and is virtually parallel to the molecular two-fold axis. At the apex of the channel, two small pockets or alcoves extend to either side. Two other residues in greenish yellow are shown: Ser 530 (at the top of the channel where it joins the alcoves, just below Tyr 385) and Arg 120 (midway up the channel). In *b*, the upper part of the cyclooxygenase channel is shown in stereo as a ribbon representation with the NSAID flurbiprofen docked into position. Tyr 385 (in light blue) is at the apex of the channel and is positioned near the edge of the haem. Ser 530 (in violet), the acetylation site for aspirin, is also at the apex of the channel; Ser 530 is not required for cyclooxygenase activity⁴⁶, but aspirin acetylation may inhibit PGHS by blocking access of arachidonic acid to the active centre. The *S*-stereoisomer of flurbiprofen (dark blue) in the channel interacts, via its carboxylate, with Arg 120 (green), thereby placing the second phenyl ring within van der Waals contact of Tyr 385. The methyl group at the asymmetric carbon C-2 is directed into a small pocket along the channel wall; docking experiments using the *R*-stereoisomer reveal a potential clash between the C-2 methyl group and Tyr 355, providing an explanation for the enzyme's stereoselectivity.



on the luminal side of the endoplasmic reticulum membrane; therefore, the bulk of the protein faces the lumen. Helices A, B, C and the beginning of helix D are amphipathic (Fig. 4*a-e*) and are situated with their hydrophobic surfaces facing outward, away from the body of the protein, while the corresponding surfaces on the adjacent monomer face the same direction (Fig. 4*e*), forming a large hydrophobic patch on the exterior of the protein. We propose that this patch anchors the dimer in the

membrane in a manner analogous to the way that surface-active peptides bind to membranes³³, with their lower surfaces interacting with the hydrophobic interior of the bilayer. The opposite sides of the helices contain a few charged residues (mainly arginines) which are well positioned to interact with phospholipid head groups. The binding surface is not deep enough to extend beyond one leaflet of the lipid bilayer (Fig. 4*e*) and thus PGHS is a monotopic membrane protein, according to Blobel's

FIG. 4 *a-d*, Schematic representations of helices A-D of the membrane-binding motif show their strongly amphipathic character. The hydrophobic outer surface of each of these helices is shown pointing downward in the figure. *e*, The PGHS dimer is seen, viewed perpendicular to the molecular two-fold axis, which is vertical in this figure. Dimyristoyl phosphatidylglycerol molecules in extended conformations⁴⁷ are included for scale and placed near the membrane-binding motif of PGHS, although exact localization of the membrane boundaries is not possible at this stage. The hydrophobic side chains on the surface of the dimer's membrane-binding motif are shown. Note how helices A-C would be roughly parallel to the membrane plane. *f*, A stereo view of the C α backbone of the EGF module from PGHS-1 (residues 34-72; in red) superimposed with the corresponding C α backbones (residues 4-45) of 16 energetically equivalent conformers (in blue) from the NMR structure of murine EGF³⁶. The intra-domain disulphide bridges are depicted by dotted lines. Differences between the two structures reflect the different lengths between the loops formed by the three disulphide bridges. For PGHS, loops I-III (using the definitions of Tulinsky and co-workers³⁷) are 12, 17 and 11 residues long, respectively; in murine EGF, they are 15, 18, 10 residues long, respectively. Least-squares superposition of the six cysteine residues (using C α , C β and S γ atoms) from the two structures gives a r.m.s. deviation of 1.3 Å for the 18 atoms. An equivalent superposition between the EGF-like modules in PGHS and factor Xa gives a r.m.s. deviation of 1.4 Å, a value similar to those found for superpositions of factor Xa and the known NMR structures of EGF and EGF-like modules³⁷. In PGHS, the N-terminal portion of the EGF-like module (residues 33-50, extending up to the turn between the two strands of the first β -sheet) and residues 58-61 create a surface that makes both intra- and inter-subunit contacts with the catalytic domains. The equivalent surface created by the EGF-like module in factor Xa is involved in A-chain interactions with the catalytic domain³⁷. The observations that (1) two PGHS monomers are connected by a large interface which includes the EGF-like module; (2) that this interface has the same orientation, relative to the catalytic domain, as the subunit interface in MPO²²; and (3) that detergent-solubilized PGHS is known to form dimers⁶ suggest that the native oligomeric state of the protein in the membrane is a dimer.

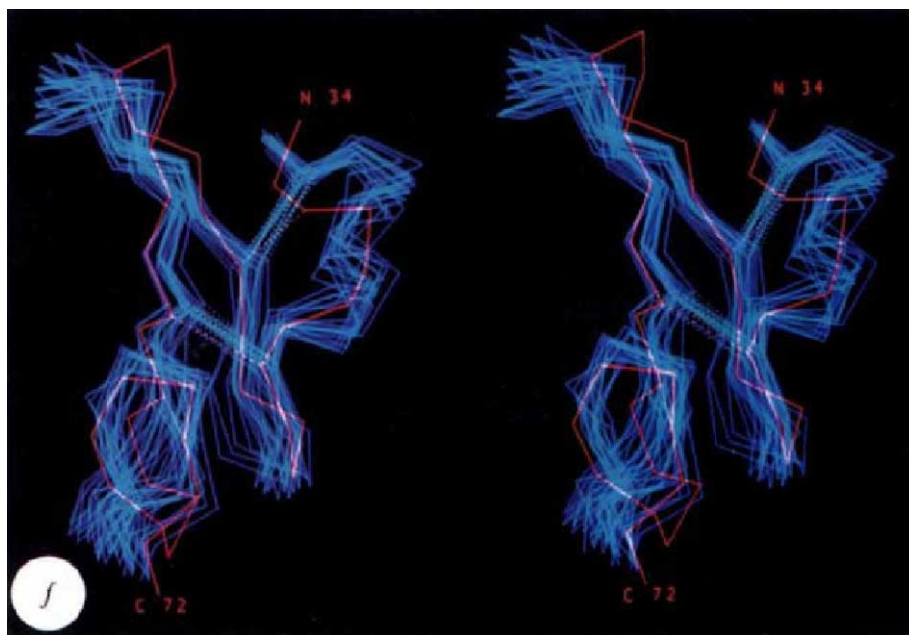
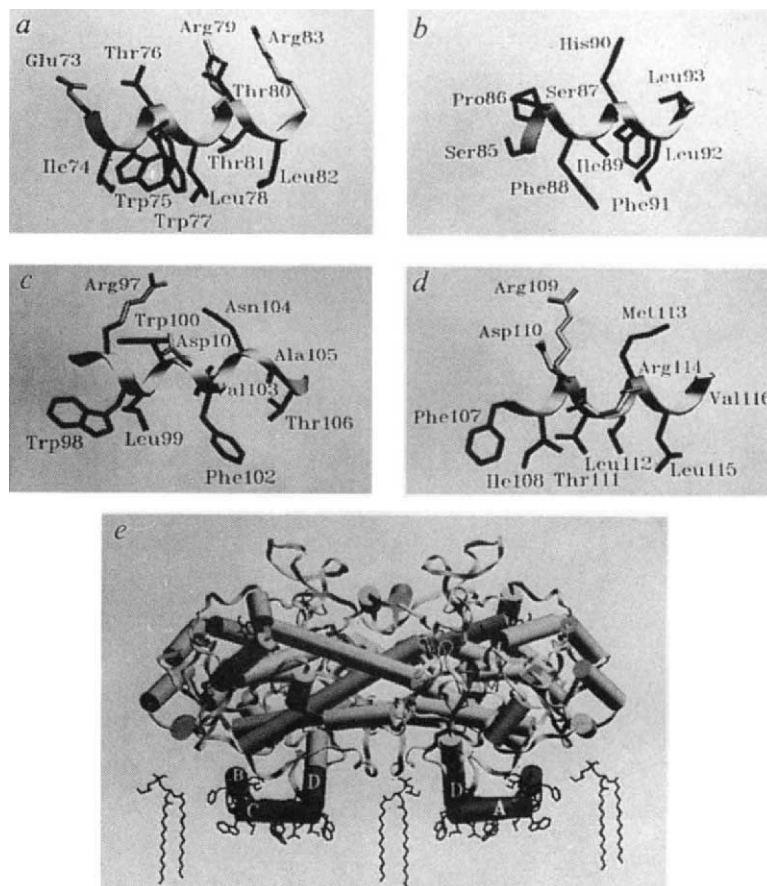


TABLE 2 Helices in PGHS-1

A*	74–82†
B	86–92
C	97–105
D	108–122
E	139–143
1	174–181
2	196–207
3	238–244
–	
5	296–319
6	325–353
–	
8	379–385
9	413–417
10	419–428
–	
11/12	445–457
13	463–469
14	478–482
15	486–495
16	503–509
17	520–535
18	553–561
19	564–569

* The nomenclature for the helices 1–19 is based on the spatial equivalence of secondary structure between canine myeloperoxidase²² and PGHS-1 but does not imply any structural or spatial correspondence between individual pairs of residues in the secondary structures. Helices identified with a letter have no equivalent in MPO.

† The accuracy of the helix boundaries is limited by the inability to identify carbonyl groups at the present resolution.

classification³⁴. This model for membrane interaction also explains how substrate effects entry into the cyclooxygenase channel. Arachidonic acid is hydrophobic, and, once released from membrane phospholipids, will presumably remain associated with the bilayer. As helices A–D form the entrance to the cyclooxygenase channel (Fig. 3a), their insertion into the mem-

brane could allow a fatty-acid molecule to gain direct access to the hydrophobic channel from the interior of the bilayer.

The EGF-like module

The EGF-like module in PGHS, corresponding to exon 3 of human PGHS³⁵, is formed by two small, two-stranded β -sheets held together by three intra-domain disulphide bonds. The module is located at the dimer interface just before the membrane-binding motif. Its structure is very similar to those of murine EGF³⁶ (Fig. 4f) and the second EGF-like module of factor Xa³⁷, as predicted by Toh³⁸. The homology between the amino-terminal domain of PGHS and EGF extends beyond the boundaries of the sequence homology usually seen in EGF-like modules¹⁷, persisting up to the Leu 78–Arg 79 dipeptide of PGHS, which falls in the middle of helix A. The equivalent residues in the murine hormone, Leu 52–Arg 53, form one of the proteolytic cleavage sites that yield mature EGF.

EGF-like modules are found in many extracellular proteins, blood clotting factors and cell-surface proteins¹⁷. Their functions are unknown, but they are thought to serve as structural building blocks that initiate or maintain protein–protein interactions. The homology between the PGHS and factor Xa EGF-like modules extends to the surface that they present towards their respective catalytic domains (Fig. 4f). The carboxy-terminal portions of the EGF-like modules, which contain the final β -sheet, face away from the catalytic domains and towards the solvent. This region in PGHS contains Asn 68, a site of *N*-linked glycosylation, which is equivalent to Arg 41 in human EGF, a residue essential for receptor binding.

In summary, the PGHS-1 structure has raised interesting questions about this protein's mechanism of insertion into the membrane, the role of the EGF-like domain, and the possible evolution of the cyclooxygenase active site within an ancestral peroxidase. The current model has provided new insights into the general features of the cyclooxygenase and peroxidase reaction mechanisms and the modes of NSAID inhibition. Crystals of the holoenzyme, grown in the absence or presence of various inhibitors, are under investigation and will provide the basis for detailed investigations of structure–function relationships in PGHS-1. □

Received 14 October; accepted 30 November 1993.

- Smith, W. L., Borgeat, P. & Fitzpatrick, F. A. in *Biochemistry of Lipids, Lipoproteins and Membranes* (eds Vance, D. E. & Vance, J.) 297–325 (Elsevier, New York, 1991).
- Marnett, L. J. *Environ. Health Perspect.* **88**, 5–12 (1990).
- Makheja, A. N. *Molec. cell. Biochem.* **111**, 137–142 (1992).
- Abramson, S. *Curr. Opin. Rheumatol.* **3**, 336–340 (1991).
- Vane, J. & Botting, R. *FASEB J.* **1**, 89–96 (1987).
- Smith, W. L. & Marnett, L. J. *Biochim. biophys. Acta* **1083**, 1–17 (1991).
- Willard, J., Lange, R. A. & Hillis, L. D. *New Engl. J. Med.* **327**, 175–181 (1992).
- Lands, W. E. M., Kulmacz, R. J. & Marshall, P. J. in *Free Radicals in Biology* Vol. VI (ed. Pryor, A. W.) 34–63 (Academic, London, 1984).
- DeWitt, D. L. *Biochim. biophys. Acta* **1083**, 121–134 (1991).
- Holtzman, M. J., Turk, J. & Shornick, L. P. *J. biol. Chem.* **267**, 21438–21445 (1992).
- Meade, E. A., Smith, W. L. & DeWitt, D. L. *J. biol. Chem.* **268**, 6610–6614 (1993).
- Otto, J. C., Dewitt, D. L. & Smith W. L., *J. biol. Chem.* **268**, 18234–18242 (1993).
- Masferrer, J. L., Seibert, K., Zweifel, B. & Needleman, P. *Proc. natn. Acad. Sci. U.S.A.* **89**, 3917–3921 (1992).
- O'Banion, M. K., Sadowski, H. B., Winn, V. & Young, D. A. *J. biol. Chem.* **266**, 23261–23267 (1991).
- Picot, D. & Garavito, R. M. in *Cytochrome P450: Biochemistry and Biophysics* (ed. Schulz, I.) 29–36 (Taylor and Francis, Philadelphia, 1989).
- Garavito, R. M. & Picot, D. *Methods: A Companion to Methods in Enzymology* **1**, 57–69 (1990).
- Campbell, I. D. & Bork, P. *Curr. Opin. struct. Biol.* **3**, 385–392 (1993).
- Marnett, L. J. & Maddipati, K. R. in *Peroxidases in Chemistry and Biology* Vol. I (eds Everse, J., Everse, K. E. & Grisham, M. B.) 293–334 (CRC Press, Boca Raton, 1991).
- Lambier, A. M., Markey, C. M., Dunford, H. B. & Marnett, L. J. *J. biol. Chem.* **260**, 14894–15000 (1985).
- Finzel, B. C., Poulos, T. L. & Kraut, J. *J. biol. Chem.* **259**, 13027–13036 (1984).
- Poulos, T. L., Edwards, S. L., Wariishi, H. & Gold, M. H. *J. biol. Chem.* **268**, 4429–4440 (1993).
- Zeng, J. & Fenna, R. E. *J. molec. Biol.* **226**, 185–207 (1992).
- Erman, J. E., Vitello, L. B., Miller, M. A. & Kraut, J. *J. Am. chem. Soc.* **114**, 6592–6593 (1992).
- Fita, I. & Rossmann, M. G. *J. molec. Biol.* **185**, 21–37 (1985).
- Hamberg, M. & Samuelsson, B. *J. biol. Chem.* **242**, 5336–5340 (1967).
- Smith, W. L., Eling, T. E., Kulmacz, R. J., Marnett, L. J. & Tsai, A. *Biochemistry* **31**, 3–7 (1992).

- Dietz, R., Nastainczyk, W. & Ruf, H. H. *Eur. J. Biochem.* **171**, 321–328 (1988).
- Karthein, R., Dietz, R., Nastainczyk, W. & Ruf, H. H. *Eur. J. Biochem.* **171**, 313–320 (1988).
- Shimokawa, T., Kulmacz, R. J., DeWitt, D. L. & Smith, W. L. *J. biol. Chem.* **265**, 20073–20076 (1990).
- Flower, R. J. *Pharmac. Rev.* **26**, 33–67 (1974).
- Mertle, J. P., Fagan, D., Mudd, J. & Needleman, P. *J. biol. Chem.* **263**, 3550–3553 (1988).
- DeWitt, D. L. & Smith, W. L. *Proc. natn. Acad. Sci. U.S.A.* **85**, 1412–1416 (1988).
- Kaiser, E. T. & Kezdy, F. A. *Rev. Biophys. biophys. Chem.* **16**, 561–581 (1987).
- Blobel, G. *Proc. natn. Acad. Sci. U.S.A.* **77**, 1496–1500 (1980).
- Yokoyama, C. & Tanabe, T. *Biochem. biophys. Res. Commun.* **165**, 888–894 (1989).
- Montelione, G. T. et al. *Biochemistry* **31**, 236–249 (1992).
- Padmanabhan, K. et al. *J. molec. Biol.* **232**, 947–966 (1993).
- Toh, H. *FEBS Lett.* **258**, 317–319 (1989).
- Zhang, K. Y. J. *Acta crystallogr. D* **49**, 213–222 (1993).
- Jones, T. A. in *Computational Crystallography* (ed. Sayre, D.) 303–317 (Oxford University Press, 1992).
- Lüthy, R., Bowie, J. U. & Eisenberg, D. *Nature* **356**, 83–85 (1992).
- Brunker, A. T. *X-PLOR Manual, Version 3.1* (Yale Univ. Press, New Haven, 1992).
- Kulmacz, R. J. *J. biol. Chem.* **264**, 14136–14144 (1989).
- Shimokawa, T. & Smith, W. L. *J. biol. Chem.* **266**, 6168–6173 (1991).
- Evans, S. V. *J. molec. Graphics* **11**, 134–138 (1993).
- Shimokawa, T. & Smith, W. L. *J. biol. Chem.* **267**, 12387–12392 (1992).
- Pascher, I., Sundell, S., Harlos, H. & Eibl, H. *Biochim. biophys. Acta* **896**, 77–88 (1987).
- Messerschmidt, A. & Pfleger, J. W. *J. appl. Crystallogr.* **20**, 306–315 (1987).
- Kabsch, W. *J. appl. Crystallogr.* **21**, 916–924 (1988).

ACKNOWLEDGEMENTS. We thank H. Einspahr and the Upjohn Company for samples of flurbiprofen, and R. Fenna and A. Tulinsky for use of the coordinates of human myeloperoxidase (Fig. 2c) and the EGF-like module from Factor Xa, respectively. The work was supported by an NIH Program Project grant and in part by a grant from the American Cancer Society. D.P. was partially supported by fellowships from Roche Research Foundation (Switzerland) and the Swiss NSF. P.J.L. was supported by a Damon Runyon–Walter Winchell Cancer Fund postdoctoral fellowship. Data collection was done at the departmental X-ray facility, established in part by the PHS/NIH Shared Instrumentation Grant. CCP4 is SERC (UK) Collaborative Computing Project No. 4, a suite of programs for protein crystallography produced at Daresbury Laboratory, Warrington, UK (1992). This work was presented in preliminary form at the American Crystallographic Association annual meeting, May 1993. Coordinates for the PGHS-1 dimer have been deposited in the Brookhaven Data Bank.

Detection of strong iron emission from quasars at redshift $z > 3$

Richard Elston*, Keith L. Thompson†
& Gary J. Hill‡

* Cerro Tololo Inter-American Observatory, Casilla 603 La Serena, Chile

† Naval Research Laboratory, Code 7210, 4555 Overlook Avenue SW, Washington DC 20375, USA

‡ McDonald Observatory University of Texas at Austin, RLM 15.308, Austin, Texas 78712, USA

QUASARS are distant, luminous objects generally thought to be powered by the accretion of gas onto a supermassive black hole¹; their spectra are characterized by broad emission lines originating from a dense region close to the central energy source¹. The best-studied spectral region in low-redshift quasars is near the $H\beta$ line at 4,861 Å (in the quasar rest frame) where there are also lines arising from singly ionized iron and doubly ionized oxygen. New technology has enabled us to detect strong iron emission in the spectra of the high-redshift ($z > 3$) quasars Q0014+813 and Q0636+680, in which these lines are redshifted to the near-infrared. The strength of this emission suggests an iron abundance (relative to hydrogen) higher than in the solar neighbourhood. This high iron abundance supports the view that quasars are located in the centres of massive galaxies. If type Ia supernovae are responsible for the iron enrichment², significant star formation must have taken place in the host galaxies at least one billion years earlier, providing a constraint on the age of the Universe at that redshift.

Permitted emission from singly ionized iron (Fe II) was first detected by Wampler and Oke in 1967 in the quasar 3C273 in the wavelength regions near 4,570 and 5,320 Å (ref. 3). The same features were then observed by Sargent in the Seyfert 1 galaxy I Zw 1 (ref. 4). Fe II has since been found in most Seyfert 1 galaxies⁵, which are also thought to be powered by gas accretion onto a black hole, and in many, but not all, low- and moderate-redshift quasars⁶⁻⁸. The presence of Fe II emission provides a challenge to emission-line models based on photoionization alone^{6,9}, and the problems are accentuated by a handful of quasars with very strong Fe II emission, relative to $H\beta$ (IRAS18508-7815 (ref. 10), IRAS07598+650 (ref. 11), PHL1092 (ref. 12), MKN231 (ref. 13)). Indeed, the great variation in the strengths of Fe II emission is in sharp contrast to the fairly uniform relative strength of other permitted lines (for example, H II, C IV and N IV), suggesting a different origin. For example, theoretical models have been proposed where the Fe II emission originates in an extended warm region (perhaps heated by soft X-rays¹⁴) at high optical depth, where hydrogen is only partly ionized and little emission occurs from other species¹⁵. In any case, the mechanism by which Fe II is excited in active galactic nuclei (AGN) remains poorly understood.

Figure 1 (top two traces) shows the near infrared K-band spectra of the luminous quasars Q0014+813 and Q0636+680, obtained using the Kitt Peak National Observatory (KPNO) cryogenic spectrometer at the $f/15$ focus of the KPNO 2.1-m telescope in December 1992. The bottom trace in Fig. 1 is the spectrum of the extreme Fe II emitting quasar IRAS07598+6508 (D. Hines, private communication). We have marked the positions of $H\beta$, [O III] 4,959 and 5,007 Å lines, and the most prominent lines of three Fe II multiplets (labelled (42), (48) and (49)) for comparison. Each multiplet consists of emission lines originating from a common upper term and ending on a common lower term, albeit with different J values.

The iron multiplet lines shown in Fig. 1 account for most of the strongest peaks in the Fe II blend to the red of $H\beta$ seen in

AGN spectra⁶. A number of these lines match peaks in the spectra of the high-redshift quasars, and the overall shapes of the spectra are similar. In particular, Q0636+680 clearly shows the 5,169 Å line of multiplet 42 at 15σ significance (where σ is one standard deviation). On the basis of these spectra, we claim detection of iron in the broad-line regions of Q0014+813 and Q0636+680. The small offset of the line peaks from their expected positions in the Q0014+813 spectrum is consistent with the expected wavelength calibration uncertainty. It is difficult to measure the overall strengths of the Fe II lines reliably because blended Fe II lines blanket the observed wavelength regions, but conservative placement of the continuum (as high as the features in the spectrum will allow) implies that the Fe II emission is strong relative to $H\beta$. Q0636+680 can probably be classified as a 'very strong iron emitter' having the intensity ratio of Fe II/ $H\beta$ significantly greater than unity, and it is possible that these two high-redshift quasars have Fe II emission strengths comparable to IRAS07598+6508. We note that all four $z \sim 2.3$ quasars studied by us¹⁶ have Fe II (4,534–4,684 Å)/ $H\beta > 1$, and would be classified as 'strong iron emitters' according to standard definitions¹⁷. In addition to the lower energy-level multiplets shown in Fig. 1, a line at $\sim 5,070$ Å is evident at $2-3\sigma$ significance in all three spectra. Although the identification of this line is unknown, we note that there are several higher energy-level Fe II transitions with large transition probabilities which might account for this line.

The relative intensity ratios of Fe II lines near $H\beta$ in low-redshift objects show some variation from object to object, but the distinctions can be traced primarily to the overall strength of iron multiplet 42 (ref. 8). Our high-redshift spectra have relative line strengths different from the low-redshift spectra, to a degree far exceeding the low-redshift variations. In particular, the three lines of multiplet 42 ($a^6S-z^6P^0$: 4,924 Å, $5/2-3/2$; 5,018 Å, $5/2-5/2$; 5,169 Å, $5/2-7/2$) normally have approximately equal strengths, but in Q0636+680, the 5,169 Å line appears stronger by a factor of more than two. The upper-level term of each multiplet is made of sub-levels with a range of J values, and collisional redistribution of these sub-levels is thought to tie their relative populations together and provide relative emission line strengths only weakly dependent on the exact physical conditions. This need not be the case. If collisional redistribution of these sub-levels is weak, mechanisms that populate the upper levels unevenly can produce a wide range of relative ratios. Fluorescent effects between Fe II lines¹⁵, or between other ions and Fe II (ref. 18), for example, might provide such excitation. Because of the consistency among the relative line ratios of low-redshift optical Fe II emitters, the hypothesis of weak collisional redistribution suggests significant differences in the conditions of the broad-line clouds for these high-redshift quasars.

Correlations of the strength of iron emission with other observed properties of quasars (for example the radio and X-ray regions of the spectrum) have been sought to elucidate the Fe II emission mechanism. As strong Fe II emission seems to be nearly ubiquitous among luminous high-redshift quasars, it is of interest to look at such properties in these objects. Most strong Fe II emitters tend to be radio-quiet quasars^{6,7}. Flat-spectrum radio-loud quasars tend to have similar Fe II emission to radio-quiet quasars, whereas steep-radio-spectrum quasars have much weaker Fe II emission^{8,19-21}. Both Q0014+813 and Q0636+680 are flat-spectrum radio sources with spectral indices of $\alpha = -0.16$ and 0.87, respectively (where $S \propto \nu^\alpha$: S is flux density and ν is frequency) (refs 22, 23), so the strong Fe II emission is not surprising. However, the only high-redshift quasar which clearly does not show Fe II emission (Q1331+170) (ref. 24) is also a flat-spectrum radio source. Although Q0014+813 seems to have been detected by the Exosat satellite with $S = 0.17 \pm 0.05 \mu\text{Jy}$ at 3 keV (ref. 25), neither source has a good X-ray slope measurement so we cannot comment on their relation to the controversial correlation between Fe II strength and soft X-ray slopes^{26,27}.

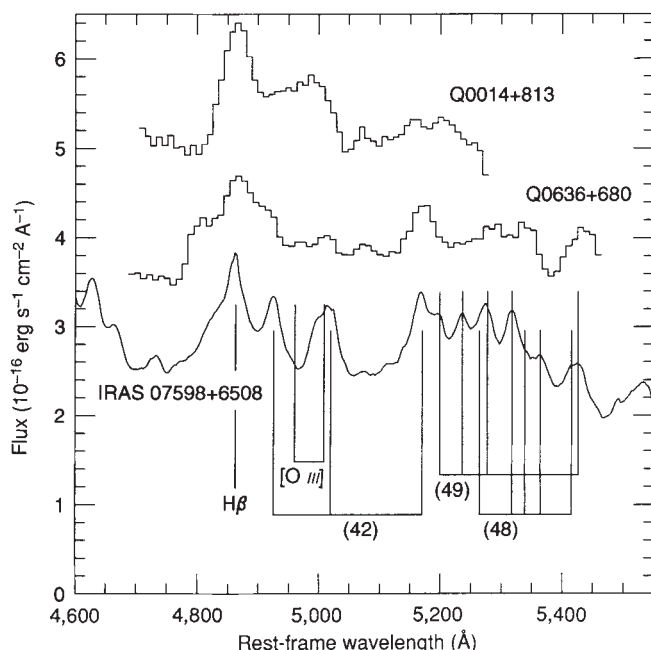


FIG. 1 Near infrared K-band spectra of Q0014 + 813 and Q0636 + 680, with the spectrum of the strong iron emitter IRAS07598 + 6508 (continuous curve) also shown for comparison. (The spectra are presented in the rest frame, and the spectrum of Q0636 + 680 is shifted up by $5 \times 10^{-17} \text{ erg s}^{-1} \text{ cm}^{-2} \text{ Å}^{-1}$ for clarity.) A grating with 150 lines per mm was used in second order giving a 2 pixel resolution of 270 (80 Å, $\sim 1,100 \text{ km s}^{-1}$) at 2.2 μm . The data were reduced and calibrated using our standard procedures¹⁶. Wavelength calibration is estimated to be accurate to <0.5 pixels. The spectra have been smoothed to 3 pixel ($1,600 \text{ km s}^{-1}$) resolution. We have adopted $z = 3.398$ for Q0014 + 813 and $z = 3.195$ for Q0636 + 680, with reference to the apparent positions of H β , [O III] 4,959 and 5,007 Å, and several important lines of Fe II multiplets 42, 48, and 49 are marked. Note that these Fe II multiplets account for many of the peaks in the Q07598 + 6508 spectrum, but the relative amplitudes are not the same for the high-redshift quasars.

We do not detect forbidden oxygen ([O III] 5,007, 4,959 Å) emission from either Q0014 + 813 or Q0636 + 680. Previous observations of the H β region of the former object at low resolution with a continuously variable filter spectrometer were interpreted as detecting [O III] (ref. 28). Our much higher resolution

measurements suggest that the broad plateau to the red of H β does not contain substantial narrow [O III] 4,959 and 5,007 Å emission, but is primarily Fe II. A blue shift of forbidden oxygen emission of $\sim 800 \text{ km s}^{-1}$ relative to H β would be required to align the 5,007 Å line with the peak to the red of H β . Although such shifts have been seen in low redshift AGN, the width of the forbidden lines necessary to account for the broad peak would be unprecedented. The spectrum of Q0636 + 680 does have small bumps of low significance at the positions of the [O III] lines, but a higher sensitivity spectrum is required before a detection of forbidden oxygen can be claimed. There is a tendency for high-redshift quasars to have weak [O III] (ref. 16). Of the seven quasars with $z > 2.0$ that have been observed at wavelengths near [O III] with sufficient sensitivity and resolution (refs 16, 22 and this work), only two have shown forbidden oxygen emission. The tendency for the luminous high-redshift quasars to have weak [O III] may be primarily due to luminosity²⁰. At low redshift, most objects with strong optical Fe II emission, compared to H β , have weak [O III] (ref. 8) the high-redshift luminous quasars seem to follow this same trend.

Because the optical Fe II lines can only be strong when the ultraviolet lines with the same upper levels are optically thick¹⁴, strong optical Fe II lines seem to require large column densities of iron. Quantitative abundance estimates are model dependent, but theoretical modelling of the spectra of low-redshift quasars suggests that the strong Fe II line emission results from iron abundances many times that of the solar neighbourhood²⁹ where type Ia supernovae are thought to provide 60% of the iron abundance^{2,30}. Thus an overabundance of iron seems to require many type Ia supernovae. The standard model for type Ia supernovae, the merger of white-dwarf binaries after a common envelope phase³¹, has an iron-enrichment timescale of the order of $10^9 \text{ yr}^{2,32,33}$. At $z = 3.4$, the age of the Universe with the Hubble constant $H_0 = 100 \text{ km s}^{-1} \text{ Mpc}^{-1}$ and deceleration parameters $q_0 = \frac{1}{2}$ would be $7 \times 10^8 \text{ yr}$, which is somewhat shorter than the timescale for iron enrichment by type Ia supernovae. Barring another source of iron enrichment, such as type II supernovae, our results imply either a smaller value of H_0 , q_0 or both. By observing higher-redshift quasars we can place more severe constraints on the cosmological parameters and the geometry and age of the Universe. The high iron abundance in these high-redshift quasars supports the view of Hamann and Ferland^{32,33} that quasars occur in highly metal-enriched environments even at the highest redshifts. Such environments would be expected at the centres of massive galaxies following their initial burst of star formation, and these observations require the initial star formation and chemical enrichment to have occurred very early in the history of the Universe. $\square$

Received 8 June; accepted 23 November 1993.

- Osterbrock, D. E. & Mathews, W. G. *Astr. Astrophys.* **24**, 171–203 (1986).
- Matteucci, F. & Greggio, L. *Astr. Astrophys.* **154**, 279–287 (1986).
- Wampler, E. J. & Oke, J. B. *Astrophys. J.* **148**, 695–704 (1967).
- Sargent, W. L. W. *Astrophys. J.* **152**, L31–L34 (1968).
- Osterbrock, D. E. *Astrophys. J.* **215**, 733–745 (1977).
- Phillips, M. M. *Astrophys. J. Suppl. Ser.* **38**, 187–203 (1978).
- Bergeron, J. & Kunth, D. *Mon. Not. R. astr. Soc.* **207**, 263–286 (1984).
- Boroson, T. A. & Green, R. F. *Astrophys. J. Suppl. Ser.* **80**, 109–135 (1992).
- Collin-Souffrin, S. et al. *Astr. Astrophys.* **166**, 27–35 (1986).
- Lipari, S., Macchetto, F. D. & Golombek, D. *Astrophys. J.* **366**, L65–L67 (1991).
- Lawrence, A. et al. *Mon. Not. R. astr. Soc.* **235**, 261–268 (1988).
- Bergeron, J. & Kunth, D. *Astr. Astrophys.* **85**, L11–L14 (1980).
- Boksenberg, A. et al. *Mon. Not. R. astr. Soc.* **178**, 451–466 (1977).
- Krolik, J. H. & Kallman, T. R. *Astrophys. J.* **324**, 714–720 (1988).
- Netzer, H. & Wills, B. J. *Astrophys. J.* **275**, 445–460 (1983).
- Hill, G. J., Thompson, K. L. & Elston, R. *Astrophys. J.* **414**, L1–L4 (1993).
- Lipari, S., Terlevich, R. & Macchetto, F. *Astrophys. J.* **406**, L451–L456 (1993).
- Penston, M. V. *Mon. Not. R. astr. Soc.* **229**, 1P–5P (1987).
- Miley, G. K. & Miller, J. S. *Astrophys. J.* **228**, L55–L58 (1979).
- Steiner, J. E. *Astrophys. J.* **250**, 469–477 (1981).
- Boroson, T. A. & Oke, J. B. *Astrophys. J.* **281**, 535–544 (1984).

- Kühr, H. et al. *Astr. J.* **86**, 854–863 (1981).
- Pauliny-Toth, I. I. K. et al. *Astr. J.* **83**, 451–474 (1978).
- Carswell, R. F. et al. *Astrophys. J.* **381**, L5–L8 (1991).
- Biermann, P. L. et al. *Astr. Astrophys. Suppl. Ser.* **96**, 339–349 (1992).
- Wilkes, B. J., Elvis, M. & McHardy, I. *Astrophys. J.* **321**, L23–L27 (1987).
- Boroson, T. A. *Astrophys. J.* **343**, L9–L12 (1989).
- Kühr, H. et al. *Astrophys. J.* **284**, L5–L8 (1984).
- Wills, B. J., Netzer, H. & Wills, D. *Astrophys. J.* **288**, 94–116 (1985).
- Wheeler, J. C., Sneden, C. & Truran, J. W. *Ann. Rev. Astr. Astrophys.* **27**, 279–349 (1989).
- Iben, I. Jr. & Tutukov, A. V. *Astrophys. J. Suppl. Ser.* **54**, 335–372 (1984).
- Hamann, F. & Ferland, G. *Astrophys. J.* **391**, L53–L57 (1992).
- Hamann, F. & Ferland, G. *Astrophys. J.* **418**, 11–27 (1993).

ACKNOWLEDGEMENTS. We thank R. Joyce and the staff of Kitt Peak for their assistance. We thank B. Wills, M. Phillips, T. Boroson, F. Hamann and G. Ferland for helpful comments, and D. Hines for the use of his IRAS07598 + 6508 spectrum. R.E. thanks the Carnegie Institution of Washington for a post-doctoral fellowship. K.L.T. thanks the National Research Council for an associateship. G.J.H. thanks the McDonald Observatory, University of Texas at Austin, for travel support. The authors are Guest Observers of the National Optical Astronomy Observatories, Kitt Peak National Observatory, which is operated by the Associated Universities for Research in Astronomy Inc., under contract with the NSF.

Superconductivity in the quaternary intermetallic compounds $\text{LnNi}_2\text{B}_2\text{C}$

R. J. Cava, H. Takagi*, H. W. Zandbergen†, J. J. Krajewski, W. F. Peck Jr, T. Siegrist, B. Batlogg, R. B. van Dover, R. J. Felder, K. Mizuhashi*, J. O. Lee*, H. Eisaki* & S. Uchida*

AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974, USA

* Department of Applied Physics, University of Tokyo, Tokyo 113, Japan

† National Centre for High Resolution Electron Microscopy, Delft University of Technology, Rotterdamsweg 137, 2628 AL Delft, The Netherlands

THE attainment of unprecedentedly high transition temperatures (T_c s) in the copper oxide superconductors illustrates how working with more complex chemical systems allows greater opportunity to balance opposing forces within a single chemical compound, leading to a better optimization of physical properties. For many desired properties, materials with optimal chemical complexity have undoubtedly not yet been found. This appears to be the case for the intermetallic superconductors, whose study has languished in recent years, and which almost never show T_c s above 15 K. These are almost all binary compounds with substitution-type additives, or, rarely, true ternary compounds such as LuRh_4B_4 ($T_c = 11.7$ K; refs 1, 2). If, as some argue (refs 3, 4), materials such as A_xC_{60} (ref. 5) and $\text{Ba}_{0.6}\text{K}_{0.4}\text{BiO}_3$ (refs 6, 7) are conventional electron-phonon superconductors with T_c s of ~ 30 K, then the absence of higher T_c s in intermetallic compounds may mean only that more complex materials have not been sufficiently explored. We have recently found superconductivity at 23 K (a T_c equal to that of the previous intermetallic record holder, Nb_3Ge ; ref. 9) in the quaternary intermetallic system yttrium-palladium-boron-carbon⁸, but we were unable to identify the superconducting phase. Here we report superconductivity at temperatures up to 16.6 K for the single-phase quaternary intermetallic compounds $\text{LnNi}_2\text{B}_2\text{C}$

(where Ln stands for Y, Tm, Er, Ho or Lu). The presence of the 3d transition metal nickel, and the layered crystal structure¹⁰ raise intriguing questions about the origin of the superconductivity, and the relatively high T_c s of these and the Y-Pd-B-C superconductor suggest that there may yet be more surprises in store.

Samples were prepared by arc-melting and annealing. The starting materials were lanthanide metal shavings or sublimed dendrites (99.9 or 99.99% pure), Ni powder (99.99%) and coarse C (99.99%) and B (99.6%) powders. Samples of 0.75 g total weight were pressed into 0.25-inch-diameter pellets, which were then arc-melted under argon on a standard water-cooled copper hearth three times, with the melted button turned over between melts. Investigation of the Lu-Ni-B-C chemical system revealed the presence of superconductivity in quaternary Ni-rich compositions. From electron microscopy and powder X-ray diffraction, in conjunction with physical property measurements, we deduced that the superconductivity was associated with a phase of tetragonal symmetry, with unit-cell dimensions $a = 3.46$ Å and $c = 10.63$ Å. This phase was obtained as very well formed single crystals directly from melts of composition $\text{LuNi}_4\text{B}_3\text{C}$, mixed with finely crystalline Ni_2B and very small quantities of a third phase. Single-crystal structure analysis⁸ revealed that the superconducting phase has the composition $\text{LuNi}_2\text{B}_2\text{C}$. Crystallized melts of this composition are not single-phase, showing that the new superconductor does not melt congruently.

Annealing the arc-melted buttons (wrapped in tantalum foil, and sealed in evacuated quartz tubes) for 12–36 hours at 1,050–1,100 °C yielded polycrystalline material with less than 2% impurity phase present. The new phase formed for the rare-earth elements Lu, Tm, Er, Ho, Y, Dy, Tb, Sm, Ce and La; other rare earths have not been investigated. Superconductivity was found not only for the non-magnetic rare-earth elements Lu and Y, but also for the magnetic rare earths Tm, Er and Ho. The other $\text{LnNi}_2\text{B}_2\text{C}$ compounds were not superconducting above 4.2 K. Superconductivity can also be observed occasionally, and in small volume fractions, in samples to which carbon has not been intentionally added—a possible explanation for a published report of superconductivity at 12 K in the Y-Ni-B system¹¹.

Figure 1 shows the temperature dependence of the magnetization measured in a field of 20 Oe in a SQUID magnetometer (Quantum Design) for single-phase polycrystalline samples of $\text{LuNi}_2\text{B}_2\text{C}$ and $\text{YNi}_2\text{B}_2\text{C}$ annealed at 1,050 °C for 16 hours. The

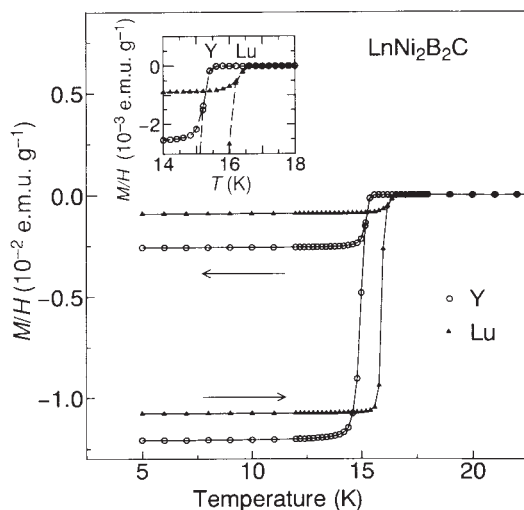


FIG. 1 Temperature-dependent magnetization (applied field $H = 20$ Oe) for $\text{LuNi}_2\text{B}_2\text{C}$ and $\text{YNi}_2\text{B}_2\text{C}$, showing the superconducting transitions. Arrows show data for warming after cooling in zero field, and data for cooling in the field.

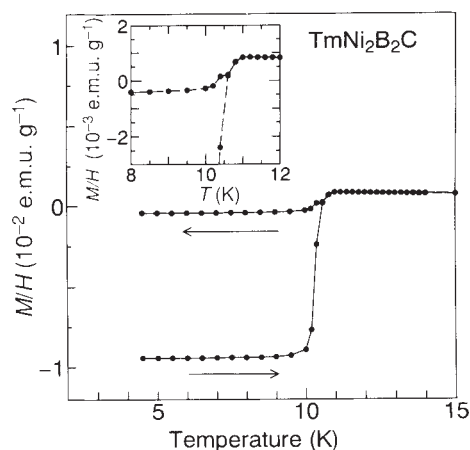


FIG. 2 Temperature-dependent magnetization ($H = 20$ Oe) for $\text{TmNi}_2\text{B}_2\text{C}$, showing the superconducting transition. Arrows show data for warming after zero-field cooling, and data for cooling in the field.

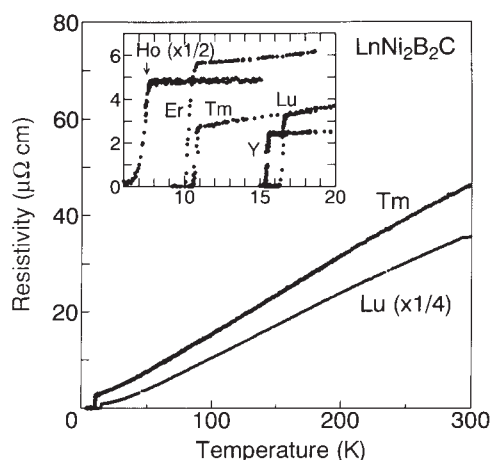


FIG. 3 Temperature-dependent resistivity between 300 and 5 K for polycrystalline samples of $\text{LuNi}_2\text{B}_2\text{C}$ and $\text{TmNi}_2\text{B}_2\text{C}$. Inset, expanded views showing the superconducting transitions for various $\text{LnNi}_2\text{B}_2\text{C}$ compounds.

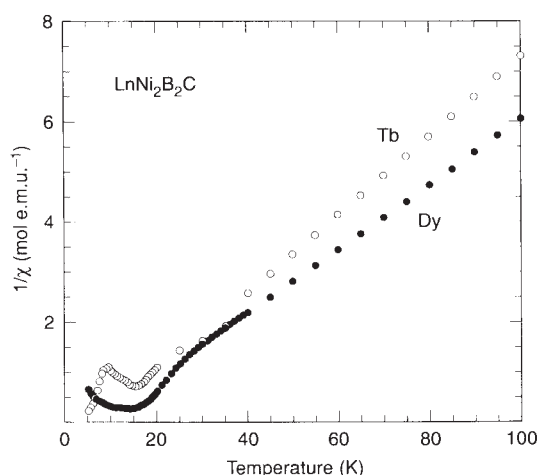


FIG. 4 Temperature-dependent magnetization for $\text{DyNi}_2\text{B}_2\text{C}$ and $\text{TbNi}_2\text{B}_2\text{C}$ ($H=100$ and 20 Oe, respectively). Magnetic transitions are seen at low temperatures. No superconductivity occurs above 4.2 K.

samples cooled in zero field show a magnetic shielding equal to approximately 100% of that expected for perfect diamagnetism. On cooling in the field, the flux expulsion values (Meissner effect) correspond to approximately 10% of that expected for perfect diamagnetism, showing that the superconductivity is a bulk effect. The existence of hysteresis between zero-field cooling and field cooling indicates that the new materials are type II superconductors. Note the sharpness of the superconducting transitions ($T_c=16.6$ K for $\text{LuNi}_2\text{B}_2\text{C}$ and 15.6 K for $\text{YNi}_2\text{B}_2\text{C}$), indicating a high degree of material uniformity.

Figure 2 shows the superconducting transitions of one variant made with a magnetic rare-earth element, $\text{TmNi}_2\text{B}_2\text{C}$. The shielding and flux expulsion values are comparable to those in Fig. 1. The T_c values are 8.0 , 10.5 and 11.0 K for $\text{HoNi}_2\text{B}_2\text{C}$, $\text{ErNi}_2\text{B}_2\text{C}$ and $\text{TmNi}_2\text{B}_2\text{C}$, respectively. The presence of the magnetic rare earths suppresses T_c , indicating that the magnetic f electrons are pair-breaking. The variation in T_c on going from $\text{LuNi}_2\text{B}_2\text{C}$ to $\text{YNi}_2\text{B}_2\text{C}$, on the other hand, is only 1 K, probably due to the change in unit-cell size. The lattice parameters change from $a=3.46$, $c=10.63$ Å for $\text{LuNi}_2\text{B}_2\text{C}$ to $a=3.53$, $c=10.57$ Å for $\text{YNi}_2\text{B}_2\text{C}$ (ref. 10). The unit cell sizes for $\text{HoNi}_2\text{B}_2\text{C}$ and $\text{YNi}_2\text{B}_2\text{C}$ are very similar¹⁰ suggesting that the 7 K difference in T_c is due to the magnetic pair-breaking in $\text{HoNi}_2\text{B}_2\text{C}$.

Figure 3 shows the temperature dependence of the resistivity between 300 and 5 K for single-phase polycrystalline samples of $\text{LuNi}_2\text{B}_2\text{C}$ and $\text{TmNi}_2\text{B}_2\text{C}$. The room-temperature normal-state resistivities are between ~ 40 and 150 $\mu\Omega$ cm for polycrystalline samples, somewhat higher than what is observed in single crystals (H.T. *et al.*, manuscript in preparation), but the resistivities just above T_c , 2 – 10 $\mu\Omega$ cm, are very low. For polycrystalline samples there are no significant differences among the resistivities for the magnetic and non-magnetic rare earths for which superconductivity is seen, suggesting at most weak coupling between the conduction electrons and the rare-earth magnetic moments. The Fig. 3 inset shows the region near T_c in detail for

various rare-earth elements. The 10 – 90% transition widths are typically ~ 0.5 K.

Along the rare-earth series, superconductivity disappears above 4.2 K for $\text{LnNi}_2\text{B}_2\text{C}$ between holmium and dysprosium. Figure 4 shows the temperature dependence of the magnetizations for $\text{DyNi}_2\text{B}_2\text{C}$ and $\text{TbNi}_2\text{B}_2\text{C}$. The present compounds appear to be similar to the LnRh_4B_4 family of superconductors, in which the break between superconducting and non-superconducting compounds for the small rare earths occurs between erbium and holmium¹². Fits to the data in Fig. 4 and to the magnetizations at higher temperatures show the rare-earth moments to be within 3% of their expected free-ion values, with weak ferromagnetic interactions. The data also suggest that the magnetic ordering transitions seen at low temperatures are rather complex.

We have shown that superconductivity occurs at high temperatures in the intermetallic compounds $\text{LnNi}_2\text{B}_2\text{C}$ —to our knowledge the first true quaternary intermetallic superconductors. The results raise many questions, one of which concerns the role of the transition metal nickel: if the $\text{Ni } 3d$ electrons are involved in the conductivity then electron–electron correlation (due to the presence of a partially filled $3d$ band) may play a significant role in the determination of the properties. The layered crystal structure is reminiscent of those of the copper oxide superconductors, but whether the materials are electronically two-dimensional remains to be seen. Most important for the materials scientist, however, is whether or not our new superconducting Pd- and Ni-based boride carbides represent the first examples of a new family of high- T_c intermetallic superconductors that will grow in number and T_c with continued investigation.

Note added in proof: After the completion of this work we became aware of another report of superconductivity in the Y–Ni–B–C chemical system¹³. We believe that the superconducting phase in their samples is $\text{YNi}_2\text{B}_2\text{C}$. □

Received 27 December; accepted 29 December 1993.

1. Johnston, D. C. *Solid St. Commun.* **24**, 699–702 (1977).
2. Matthias, B. T., Corenzwit, E., Vandenberg, J. M. & Barz, H. E. *Proc. natn. Acad. Sci. U.S.A.* **74**, 1334–1335 (1977).
3. Hinks, D. G., Richards, D. R., Dabrowski, B., Marx, D. T. & Mitchell, A. W., *Nature* **335**, 419–421 (1988).
4. Schlüter, M., Lanoo, M., Needels, M., Baraff, G. A. & Tomanek, D. *J. Phys. Chem. Solids* **53**, 1473–1485 (1992).
5. Hebard, A. F. *et al. Nature* **350**, 600–601 (1991).
6. Mattheiss, L. F., Gyorgy, E. M. & Johnson, D. W. Jr *Phys. Rev.* **B37**, 3745–3746 (1988).

7. Cava, R. J. *et al. Nature* **332**, 814–816 (1988).
8. Cava, R. J. *et al. Nature* **367**, 146–148 (1994).
9. Flükiger, R. in *Concise Encyclopedia of Magnetic and Superconducting Materials* (ed. Evetts, J.) 1–15 (Pergamon, Oxford, 1992).
10. Siegrist, T., Zandbergen, H., Cava, R. J., Krajewski, J. J. & Peck, W. F. Jr *Nature* **367**, 254–256 (1994).
11. Mazumdar, C. *et al. Solid St. Commun.* **87**, 413–416 (1993).
12. Johnston, D. C. & Brown, H. F. in *Superconductivity in Ternary Compounds III* (eds Maple, M. B. & Fischer, O.) 11–51 (Springer, Berlin, 1982).
13. Nagarajan, R. *et al. Phys. Rev. Lett.* **72**, 274–277 (1994).

The crystal structure of superconducting $\text{LuNi}_2\text{B}_2\text{C}$ and the related phase LuNiBC

T. Siegrist, H. W. Zandbergen*, R. J. Cava, J. J. Krajewski & W. F. Peck Jr

AT&T Bell Laboratories, 600 Mountain Avenue, Murray Hill, New Jersey 07974, USA

* National Centre for High Resolution Electron Microscopy, Delft University of Technology, Rotterdamsweg 137, 2628 AL Delft, The Netherlands

SUPERCONDUCTING intermetallic compounds with relatively high transition temperature (T_c) have been found in the systems Ln-Tr-B-C , where Ln is a lanthanide element (Y, Ho–Lu) and Tr a transition metal (Pd or Ni)^{1,2}. In the nickel-bearing system, the superconducting phase has the formula $\text{LuNi}_2\text{B}_2\text{C}$ (ref. 2). Although many structures are known for ternary carbides and borides, only a few have been reported for quaternary borocarbides³. Here we describe two new phases that form in the Lu–Ni–B–C system: $\text{LuNi}_2\text{B}_2\text{C}$, a superconductor with a T_c of 16.6 K (ref. 2), and LuNiBC , a closely related non-superconducting phase. The superconductor is a variant on the layered ThCr_2Si_2 -type structure⁴, with additional carbon in the Lu plane. LuNiBC is derived from $\text{LuNi}_2\text{B}_2\text{C}$ by adding another layer of Lu–C, producing a NaCl-type intergrowth. A new homologous series of intermetallic phases may be derivable from the simple building principles of these two structures.

Small single crystals of $\text{LuNi}_2\text{B}_2\text{C}$ were grown in an arc furnace from melts with starting composition of $\text{LuNi}_4\text{B}_3\text{C}$, and small single crystals of LuNiBC from melts with starting compo-

sition $\text{LuNi}_2\text{B}_3\text{C}$. The crystals grew as thin platelets at the surface of the melted buttons and could be cleaved off. Small thin fragments were then mounted on an Enraf–Nonius single-crystal diffractometer using graphite-monochromatized $\text{Mo K}\alpha$ radiation. Data were collected in the Ω -scan mode up to $2\theta = 75^\circ$ at an ambient temperature of 23°C . Accurate lattice parameters were obtained by determining the absolute 2θ values at high angles of at least 35 reflections⁵. A gaussian integration absorption correction was applied to all the intensity data measured. The structures were refined using the NRCVAX program package⁶, running on a Sun workstation. Crystallographic data for the two phases are given in Table 1, with atomic positions and isotropic thermal parameters in Table 2. Figure 1 shows the two new structures. The first one, $\text{LuNi}_2\text{B}_2\text{C}$ (Fig. 1a), can be regarded as a 'filled' variant of the tetragonal body-centred ThCr_2Si_2 -type structure, a structure found for a variety of rare-earth transition-metal borides⁴. The nickel boride framework of the structure is modified by the insertion of a carbon atom in the lanthanide layer, directly between the boron atoms forming the cage of the lanthanide. This modification is possible while maintaining the overall structure because of the short boron–carbon distance of only $1.47\text{ \AA}$, resulting in the expansion of the boron–boron distance to $2.94\text{ \AA}$, as compared with $2.28\text{ \AA}$ in the related ThCr_2Si_2 -type compound YCo_2B_2 (ref. 3). The addition of the carbon to the lanthanide plane manifests itself (comparing YCo_2B_2 and $\text{LuNi}_2\text{B}_2\text{C}$) in the expansion of the c -axis lattice parameter from 9.358 to $10.631\text{ \AA}$, and the contraction of the a axis from 3.561 to $3.464\text{ \AA}$. Similar lattice expansions are observed in other intermetallic compounds, notably in hydrogen storage systems⁷. The quite substantial contraction of the a axis, however, clearly indicates a strong chemical bond between the lanthanide and the carbon atoms. The interatomic distances found in the new structures are comparable to the expected values: for $\text{LuNi}_2\text{B}_2\text{C}$ and LuNiBC , the Lu–C distances are

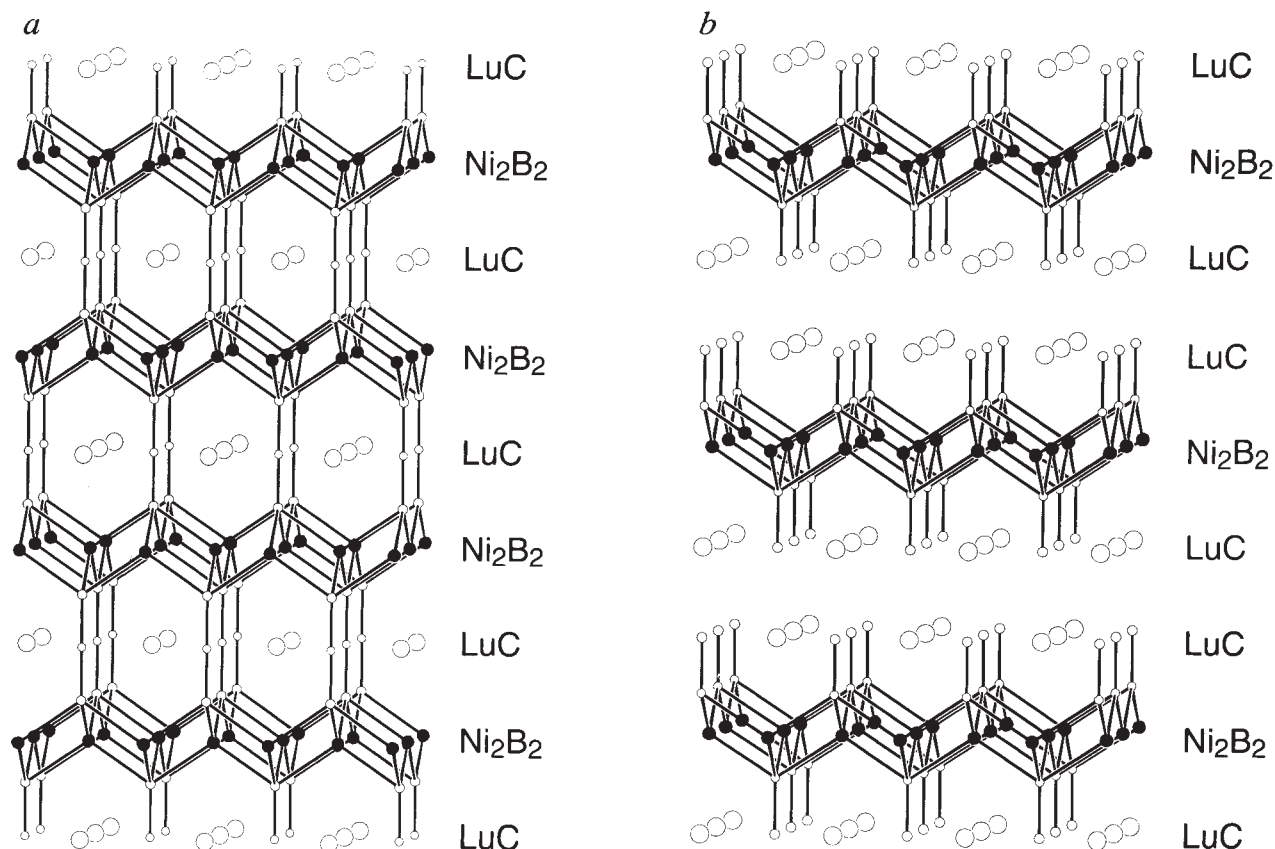


FIG. 1 The crystal structures of $\text{LuNi}_2\text{B}_2\text{C}$ (a) and LuNiBC (b). Filled circles are Ni atoms, the largest open circles are Lu atoms. The Ni–B–C framework is indicated by the bonds between the atoms.

TABLE 1 Crystallographic data for $\text{LuNi}_2\text{B}_2\text{C}$ and LuNiBC

Formula	$\text{LuNi}_2\text{B}_2\text{C}$	LuNiBC
Space group	$I4/mmm$	$P4/nmm$
Crystal size (mm)	$0.10 \times 0.05 \times 0.02$	$0.17 \times 0.08 \times 0.01$
a (Å) (at 2.3 °C)	3.4639 (1)	3.4985 (3)
c (Å)	10.6313 (4)	7.7556 (7)
v (Å ³)	127.56 (1)	94.93 (2)
Z (formula units per unit cell)	2	2
Calculated density (g cm ⁻³)	8.488	8.974
μ (cm ⁻¹)	528.3	612.7
Reflections measured	434	2079
Independent reflections	125	180
Observed reflections*	124	179
Refined parameters	11	13
R	0.028	0.028
R_w †	0.037	0.033
Extinction length (μm)	0.22 (3)	0.26 (3)

* $I > 2.5\sigma(I)$.† $w = 1/\sigma^2(F_{\text{obs}})$, where F_{obs} is the observed structure factor.

2.449 and 2.475 Å, the Lu–B distances are 2.855 and 2.867 Å, the Ni–B distances 2.10 and 2.11 Å, the Ni–Ni distances (in the Ni plane) are 2.449 and 2.474 Å, and the B–C distances are 1.47 and 1.52 Å, respectively. The two structures show a remarkable similarity in their bond lengths, indicating equivalent chemical bonding behaviour.

The new $\text{LuNi}_2\text{B}_2\text{C}$ structure, even though it is a three-dimensionally connected framework, may be viewed as a layered system, reminiscent of the high- T_c oxide superconductors. As can be seen in Fig. 1a, LuC NaCl-type layers alternate with Ni_2B_2 layers, with a stoichiometry of 1:1. The Ni_2B_2 layers are two-dimensional networks of the inverse PbO-type, with nickel being tetrahedrally coordinated by four boron atoms. The Ni_2B_2 layers contain a square-planar Ni_2 array sandwiched between the boron planes. The Ni–Ni in-plane distances of 2.45 Å are slightly shorter than those found in metallic nickel (2.50 Å), suggesting strong metal–metal bonding. The NiB_4 tetrahedra share edges to form the extended framework, with the B–Ni–B bond angles of 108.75° and 110.9° for $\text{LuNi}_2\text{B}_2\text{C}$ and 108.05° and 112.4° for LuNiBC . These angles clearly show that the Ni_2B_2 layers are rigid and are barely distorted. As a consequence, the B–Ni–B angles are also close to the tetrahedral angle, 110.9° for $\text{LuNi}_2\text{B}_2\text{C}$ and 112.4° for LuNiBC . The layers connect, not directly as in the ThCr_2Si_2 -type structure, but by way of the short boron–carbon bonds. The Lu atom is located in the cavity formed by this network, but can be chemically viewed as part of an NaCl-type layer. This is even more pronounced in LuNiBC , where now two LuC layers are stacked in NaCl-type fashion, as shown in Fig. 1b. Insertion of more LuC layers gives rise to a homologous series $(\text{LuC})_n(\text{Ni}_2\text{B}_2)$, with the two observed structures representing the $n=1$ and $n=2$ members of the series. A corresponding homologous series can be obtained by adding NiB layers, forming the compounds $(\text{LuC})(\text{NiB})_{2+n}$, and a combination of the two gives rise to $(\text{LuC})_m(\text{NiB})_n$. The $\text{LuNi}_2\text{B}_2\text{C}$ -type structure readily accepts other rare-earth ions, even up to the large lanthanum ion. We expected to see the regular lanthanide contraction according to Vegard's law^{8,9} and indeed the a axis does contract as the size of the lanthanide ion decreases (Fig. 2). The c axis, however, follows the opposite trend (Fig. 2), and as a result, the overall contraction in volume is small: from 142.5 Å³ for La to 127.6 Å³ for Lu, only an 11% decrease. With the exception of cerium, the lattice constants depend linearly on ionic radius. The deviation from this dependence for cerium indicates an intermediate or mixed valence state of approximately $\text{Ce}^{3.75+}$, revealing hybridization of the Ce 4f orbital with the broader electronic bands from the rest of the structure.

TABLE 2 Atomic positions and isotropic temperature factors for $\text{LuNi}_2\text{B}_2\text{C}$ and LuNiBC

$\text{LuNi}_2\text{B}_2\text{C}$	x	y	z	B_{iso} (Å ²)
Lu	0	0	0	0.45 (3)
Ni	1/2	0	1/4	0.40 (4)
B	0	0	0.3621 (18)	0.6 (4)
C	1/2	1/2	0	0.6 (5)

LuNiBC (origin choice 2)	x	y	z	B_{iso} (Å ²)
Lu	1/4	1/4	0.16202 (9)	0.30 (2)
Ni	3/4	1/4	1/2	0.29 (4)
B	3/4	3/4	0.1523 (21)	0.6 (3)
C	3/4	3/4	0.3489 (23)	0.6 (3)

Thin specimens for electron microscopy were obtained by crushing the melted buttons under ethanol and mounting the fragments on a copper grid covered with a holey carbon film. Electron microscopy was performed using a Philips CM30ST electron microscope equipped with a field emission gun operating at 300 kV, giving a point resolution of 0.2 nm and an information limit of ~0.12 nm.

Electron diffraction performed on a number of crystals of $\text{LuNi}_2\text{B}_2\text{C}$ occasionally showed streaking along the c^* direction, which high-resolution electron microscopy showed to be caused by the intergrowth of other phases of composition $(\text{LuC})_m(\text{NiB})_n$ ($m, n \neq 1$). Apart from this streaking, no diffuse scattering of extra reflections could be observed, indicating that the simple structure as determined by X-ray diffraction is adopted without any other long- or short-range superstructure. In contrast with the high fraction of stacking defects in the bismuth- and thallium-based oxide superconductors¹⁰, planar defects or intergrowths in $\text{LuNi}_2\text{B}_2\text{C}$ and LuNiBC are rare.

The new boride–carbide superconductor $\text{LuNi}_2\text{B}_2\text{C}$ has a layer-like crystal structure of a previously unknown type.

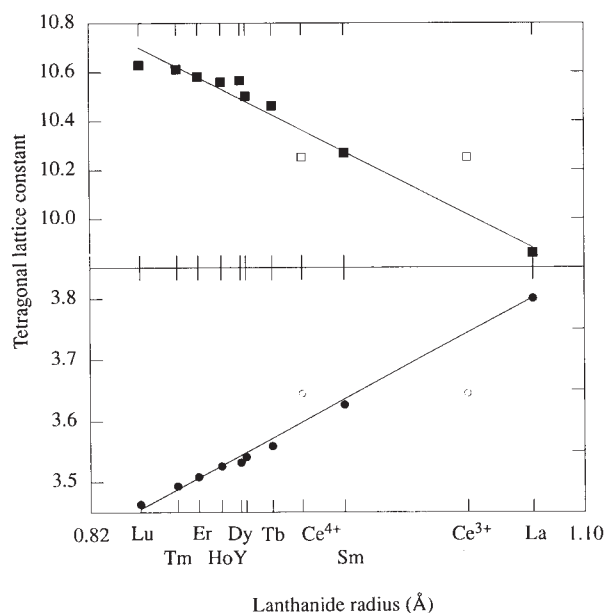


FIG. 2 Tetragonal lattice constants a (bottom) and c (top) for $\text{LuNi}_2\text{B}_2\text{C}$ with Ln represented by La, Ce, Sm, Tb, Dy, Y, Ho, Er, Tm and Lu. Ionic radii from refs 8 and 9; values are shown for both Ce^{3+} and Ce^{4+} (open symbols). The solid lines are guides to the eye.

Although a quaternary compound, the crystal structure is simple. It is one member of a homologous series built from Ni_2B_2 and LuC layers. The fact that the structure type forms for rare-earth elements ranging from the largest to the smallest shows that it is quite robust: the changing c/a ratio with lanthanide radius may reflect a small structural rearrangement to accommodate the variation in lanthanide size. Aside from the superconducting properties, interesting electronic and magnetic properties can be expected for different combinations of rare-earth and transition-metal elements, such as are found in the ThCr_2Si_2 family of compounds¹. □

Received 27 December; accepted 29 December 1993.

1. Cava, R. J. et al. *Nature* **367**, 146–148 (1994).
2. Cava, R. J. et al. *Nature* **367**, 252–253 (1994).
3. Villars, P. & Calvert, L. D. *Pearson's Handbook of Crystallographic Data for Intermetallic Phases* Vols 1–4 (ASM International, Materials Park, 1991).
4. Parthé, E. & Chabot, B. in *Handbook on the Physics and Chemistry of Rare Earths* Vol. 6 Ch. 48 (eds Gschneidner, K. A. & Eyring, L.) 113–334 (North-Holland, Amsterdam, 1984).
5. LePage, Y., White, P. S. & Gabe, E. J. Poster PA23 presented at *Am. crystallogr. Assoc. a. Meet.*, Hamilton, Canada (AIP, New York, 1986).
6. Gabe, E. J., LePage, Y., Charland, J.-P., Lee, F. L. & White, P. S. *J. appl. Crystallogr.* **22**, 348–387 (1989).
7. Buschow, K. H. J. in *Handbook on the Physics and Chemistry of Rare Earths* Vol. 6 Ch. 47 (eds Gschneidner, K. A. & Eyring, L.) 1–111 (North-Holland, Amsterdam, 1984).
8. Shannon, R. D. & Prewitt, C. T. *Acta crystallogr.* **B26**, 1046–1048 (1970).
9. Iandelli, A. & Palenzona, A. in *Handbook on the Physics and Chemistry of Rare Earths* Vol. 2 Ch. 13 (eds Gschneidner, K. A. & Eyring, L.) 1–53 (North-Holland, Amsterdam, 1984).
10. Zandbergen, H. W. et al. *Nature* **332**, 620–623 (1988).
11. Szytula, A. & Leciejewicz, J. in *Handbook on the Physics and Chemistry of Rare Earths* Vol. 12 Ch. 83 (eds Gschneidner, K. A. & Eyring, L.) 133–211 (North-Holland, Amsterdam, 1989).

ACKNOWLEDGEMENTS. We thank D. W. Murphy for valuable discussions.

Probing the interior of fullerenes by ^3He NMR spectroscopy of endohedral $^3\text{He}@C_{60}$ and $^3\text{He}@C_{70}$

Martin Saunders*, Hugo A. Jiménez-Vázquez*, R. James Cross*, Stanley Mroczkowski†, Darón I. Freedberg† & Frank A. L. Anet‡

* Department of Chemistry, † Applied Physics Department, Yale University, New Haven, Connecticut 06520, USA

‡ Department of Chemistry and Biochemistry, University of California, Los Angeles, California 90024, USA

FULLERENES have internal cavities large enough to encapsulate atoms^{1,2}. Recently, noble-gas atoms were introduced into about one in a million fullerene molecules³. It is now possible to achieve far greater yields of these noble-gas endohedral compounds⁴. Because ^3He has a spin of 1/2 and is an excellent NMR nucleus^{5,6}, it can be used as a probe for the magnetic shielding environment inside the fullerene cavity. This environment should reflect possible ring currents and hence the aromaticity in fullerenes^{7–15}, an issue that measurements of magnetic susceptibility^{16,17} have not completely resolved. Here we present ^3He NMR spectra of the endohedral compounds $^3\text{He}@C_{60}$ and $^3\text{He}@C_{70}$ (the @ symbol denotes a compound that is endohedral). We find that the ^3He nuclei encapsulated in C_{60} and C_{70} are shielded by 6 and 29 parts per million respectively, relative to free ^3He . These shieldings are unexpectedly large, indicating significant diamagnetic ring currents in C_{60} and very large ones in C_{70} . Our results also show that, because of its small size and inertness, helium can serve as a useful probe of magnetic molecular properties. In addition, our work represents the first ^3He NMR spectra of stable helium compounds.

The major improvement in introducing noble gases into C_{60} and C_{70} consists of employing high helium pressures in a modification of our earlier method³; this modified technique has been used here. A small quantity (50–100 mg) of a C_{60}/C_{70} mixture (Texas Fullerenes, Houston) was heated for 5 h at 600 °C in ^3He at a pressure of ~2,500 atm, then worked up by extraction with toluene or carbon disulphide (work done at Yale). We have previously presented a mechanism to account for our observation of the incorporation of noble gas atoms into the fullerene cavity³. We proposed that one or more bonds in the fullerene can be broken reversibly to temporarily create 'windows' large enough to allow the entry of the atoms.

For NMR spectroscopy (work done at UCLA), the ^3He -treated fullerene sample was dissolved in 1-methylnaphthalene containing chromium(III) acetylacetonate as a relaxation agent¹⁸, because a very long T_1 (spin-lattice relaxation time) was expected for ^3He inside the fullerene molecules. In most experiments, ^3He was bubbled through the solution to give a reference signal. Figure 1 shows the ^3He NMR spectra of two different treated-fullerene samples. Because the samples contain known different relative concentrations (3:1 in Fig. 1a and 1:1 in Fig. 1b) of C_{60} and C_{70} , as measured by ^{13}C NMR, an unambiguous assignment of the ^3He signals from the C_{60} and C_{70} endohedral ^3He compounds can be made, as shown in the figure.

The chemical shift of $^3\text{He}@C_{60}$ is -6.3 ± 0.15 p.p.m., that is, the ^3He inside the fullerene cage is shielded ('shifted upfield') with respect to dissolved free ^3He by ~6 p.p.m. In $^3\text{He}@C_{70}$ the chemical shift is -28.8 ± 0.2 p.p.m., and the shielding is more than four times greater than in the C_{60} compound, a truly dramatic difference. These chemical shifts remain within the limits given when the solvent is changed from 1-methylnaphthalene to 1-methylnaphthalene/ CS_2 (1:1) or to pure CS_2 , or when the temperature of the CS_2 solution is changed from room temperature to -20 °C. The same chemical shifts are obtained when gaseous helium is the reference, provided that a correction is made for the magnetic susceptibility of the solvent. We have also measured ^{13}C NMR spectra of the above samples and these show all the known peaks of C_{60} and C_{70} as well as the solvent resonances. Integration of ^{13}C and ^3He spectra taken under non-saturating conditions show that ~0.1% of fullerene molecules contain helium, and that the labelling of the C_{60} and C_{70} molecules are similar.

Before evaluating the ^3He chemical shifts, it is useful to discuss the region of the fullerene cage that is magnetically sensed by the endohedral ^3He atom. The potential-energy functions of endohedral noble gases in C_{60} and C_{70} have been studied by quantum-mechanical and molecular-mechanic (empirical force-fields) methods^{19–23}. For both helium and neon, the potential-energy function is essentially spherically symmetric with a reported minimum at the centre of the cage, except for C_{70} , where a shallow maximum or a flat curve is found at the centre in the direction of the long axis of the molecule²². The vibrational frequency of helium inside C_{60} has been calculated²³ to be 70 cm^{-1} . From these data we estimate that the ^3He atoms senses the magnetic field only inside a sphere of $<1\text{ Å}$ diameter at the centre of the C_{60} cage. The situation is probably similar in $^3\text{He}@C_{70}$, but with the spherical region elongated to a small ellipsoid mimicking the shape of that fullerene. The ^3He is essentially a probe for measuring the magnetic field at the centre of the fullerene cage, averaged by tumbling of the molecule. A 'view' of the otherwise highly inaccessible interior of the fullerene molecule can thus be obtained.

The shieldings of ^3He in $^3\text{He}@C_{60}$ and $^3\text{He}@C_{70}$ can be compared with Haddon's predictions^{7, 11} that C_{60} should have a very small (and perhaps paramagnetic) ring current and that a probe at the centre of the molecule should be much less shielded than in C_{70} . This is because the latter fullerene is calculated to have a significant diamagnetic ring current, albeit reduced as compared to graphite. Although there is qualitative agreement, Haddon's calculations produce a chemical shift of only a few

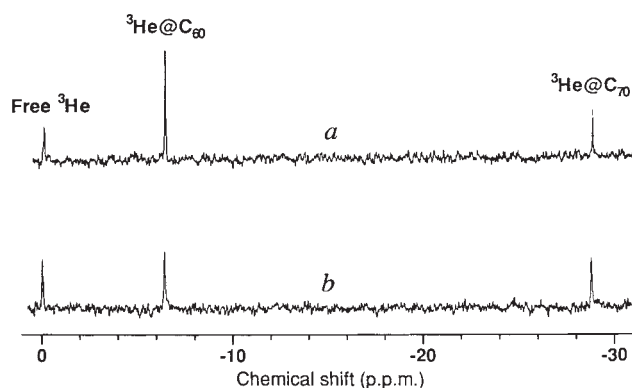


FIG. 1 ^3He NMR spectra (381 MHz) of endohedral fullerene compounds in 1-methylnaphthalene solutions containing a small amount of added dissolved (free) ^3He and 2–5 mg of chromium(III) acetylacetonate for relaxation purposes. (The chemical shift is with respect to the dissolved ^3He .) a, Soluble product (20 mg) obtained from a commercial mixture of C_{60} and C_{70} treated with ^3He ($\sim 2,500$ atm) at 600°C for 5 h. b, As a, but with the reaction product chromatographed to give a sample (15 mg) enriched in C_{70} .

METHODS. The spectra were obtained with a 5 mm proton probe modified by the addition of a 1 pF capacitor between the coil leads so that tuning to the ^3He resonance frequency at 11.7 T could be achieved. Other spectra, discussed in the text, were obtained with a 10 mm probe, where the proton decoupling coil was similarly modified. The required 381 MHz frequency was obtained either directly (Bruker ARX-500), or indirectly, by a frequency step-down and step-up procedure from the spectrometer ^1H (500 MHz) output (Bruker AM-500); in both cases a tuned power amplifier (5 to 10 W) was used to provide a 90° pulse length of ~ 20 μs . A tuned low-noise preamplifier was added between the probe and the spectrometer. The lock substance (deuterated acetonitrile or dimethyl sulphoxide) was contained in a 1 mm capillary tube centred in the sample tube. The spectra were taken at room temperature with a pulse length of 12–15 μs and an acquisition time of 0.2 s (no relaxation delay) in a total time of ~ 2 h. An exponential broadening function corresponding to 12 Hz of line broadening was used to process the free induction decay (FID) data.

p.p.m. for ^3He in $^3\text{He}@\text{C}_{60}$ and underestimates the experimental value of -6 p.p.m. This is perhaps not surprising as Haddon's theoretical calculations are based upon the London approximation and their major virtue is simplicity and qualitative rather than quantitative accuracy. *Ab initio* quantum-mechanical calculations suggest that C_{60} may have significant ring currents¹², but the interpretation of these calculations has been both disputed⁹ and supported¹³. Unfortunately, *ab initio* calculations of shieldings, especially in large molecules, are difficult^{12,14,15}. Stimulated by our present findings, J. Cioslowski (personal communication) has carried out *ab initio* (GIAO CPHF) calculations on $^3\text{He}@\text{C}_{60}$ and finds a ^3He shielding of 8.7 p.p.m.

Experimentally, the magnetic susceptibility of C_{60} has been found to be very small ($\chi_g = -0.35 \times 10^{-6}$ and -0.36×10^{-6} in c.g.s. units) compared to graphite (average $\chi_g = -7.3$) and ordinary polycyclic aromatic molecules (average $\chi_g = -0.7$ to -0.8)^{16,17}. The susceptibility of C_{70} , which has a lower fraction of five-membered rings than does C_{60} , is given experimentally^{16,17} as average $\chi_g = -0.67 \times 10^{-6}$ and -0.59×10^{-6} , which is significantly greater than that of C_{60} , but still somewhat low compared to normal aromatic molecules^{16,17}. The very different shieldings of ^3He in $^3\text{He}@\text{C}_{60}$ and $^3\text{He}@\text{C}_{70}$ found in the present work clearly do not simply reflect the magnetic susceptibilities of the fullerenes. Significantly, the proton shifts in various carbene adducts to C_{60} show that protons above a six-membered ring are shielded, in contrast to those above a five-membered ring, which are deshielded^{24,25}. Although these

adducts lack the symmetry and exact electronic structure of C_{60} , the results are in agreement with Haddon's predictions that six- and five-membered rings have diamagnetic and paramagnetic ring currents respectively, which more or less cancel, depending on the fullerene^{10,11}. A shielding (diamagnetic) contribution of ~ 4.5 p.p.m. from each six-membered ring and a deshielding (paramagnetic) contribution of ~ -7 p.p.m. from each five-membered ring would fit our experimental shielding data, but it remains to be seen whether this dissection has any general usefulness. In Cram's carcerands²⁶, which have cavities lined with eight isolated benzene rings, guest (endohedral) molecules have protons that are shielded as much as 5.3 p.p.m. from their normal chemical shifts, a value close to that found in $^3\text{He}@\text{C}_{60}$. It seems clear that C_{70} has a relatively large net diamagnetic ring current and should be classified as being aromatic. In contrast, the aromaticity of C_{60} is rather less well-defined.

All of the many reported derivatives of C_{60} and C_{70} can, in principle, be prepared from the corresponding fullerene labelled with ^3He . A strong analogy with isotopic labelling can be made, as the endohedral ^3He atom should not change the qualitative chemistry of the fullerenes. Only a slight perturbation of rates and equilibria are expected, as found in isotope-effect studies. Each of these labelled compounds should show a single peak in its ^3He NMR spectrum. Because these chemical shifts are clearly very sensitive to structure, it is unlikely that the peaks of different compounds will be accidentally coincident. The areas of these peaks should accurately yield the relative concentrations of the different compounds present. Any release of helium through mechanisms which might temporarily open 'windows' in the fullerene structure should also be readily followed from intensity changes in the ^3He NMR spectrum. Of special interest with respect to aromaticity and ring currents are the diamagnetic (presumed) hexa-anions of C_{60} and C_{70} (refs 11, 27, 28). It should be possible to label the larger fullerenes²⁹, C_n , where $n > 70$, with ^3He and to measure the shielding present inside these large cages by the methods used here. Thus, ^3He NMR spectroscopy should be a potentially powerful aid in following, as well as interpreting, the chemistry of the fullerenes and its utility will be greatly enhanced if the concentration of the labelled species could (as appears quite likely) be increased above the present value of $\sim 0.1\%$. Even at the present low concentration, the ^3He peak for $^3\text{He}@\text{C}_{60}$ becomes visible after a few minutes of data acquisition. The present work shows that ^3He NMR in chemistry is not restricted to its one previous application, namely, an investigation of solvent effects on dissolved helium gas³, where the chemical shift range was less than 1 p.p.m.

Finally, it may be appropriate to recall the remark made in 1985 by Kroto *et al.*³⁰ in their original paper on buckminsterfullerene that "the chemical shift in the NMR of a central atom should be remarkable because of ring currents". $\square$

Received 28 September; accepted 23 November 1993.

- Heath, J. R. *et al.* *J. Am. chem. Soc.* **107**, 7779–7780 (1985).
- Bethune, D. S., Johnson, R. D., Salem, J. R., de Vries, M. S. & Yannoni, C. S. *Nature* **366**, 123–128 (1993).
- Saunders, M., Jiménez-Vázquez, H. A., Cross, R. J. & Poreda, R. J. *Science* **259**, 1428–1430 (1993).
- Saunders, M. *et al.* *J. Am. Chem. Soc.* (in the press).
- Schrobilgen, G. J. in *NMR and the Periodic Table* (eds Harris, R. K. & Mann, B. E.) 439–454 (Academic, New York, 1976).
- Seydoux, R., Diehl, P., Nazitov, R. K. & Jokisaari, J. J. *magn. Reson.* **101**, 78–83 (1993).
- Elser, V. & Haddon, R. C. *Nature* **325**, 792–794 (1987).
- Elser, V. & Haddon, R. C. *Phys. Rev. A* **36**, 4579–4584 (1987).
- Haddon, R. C. & Elser, V. *Chem. Phys. Lett.* **169**, 362–364 (1990).
- Pasquarello, A., Schlüter, M. & Haddon, R. C. *Science* **257**, 1660–1661 (1992).
- Pasquarello, A., Schlüter, M. & Haddon, R. C. *Phys. Review A* **47**, 1783–1789 (1993).
- Fowler, P. W., Lazzarotti, P. & Zanasi, R. *Chem. Phys. Lett.* **165**, 79–86 (1990).
- Schmalz, T. G. *Chem. Phys. Lett.* **175**, 3–5 (1990).
- Fowler, P. W., Lazzarotti, P., Malagoli, M. & Zanasi, R. *J. phys. Chem.* **95**, 6404–6405 (1991).
- Fowler, P. W., Lazzarotti, P., Malagoli, M. & Zanasi, R. *Chem. Phys. Lett.* **179**, 174–180 (1991).
- Haddon, R. C. *et al.* *Nature* **350**, 46–47 (1991).
- Ruoff, R. S. *et al.* *J. phys. Chem.* **95**, 3457–3459 (1991).
- Martin, M. L., Delpuech, J.-J. & Martin, G. J. *Practical NMR Spectroscopy* 404–405 (Heyden, London, 1980).

19. Cioslowski, J. & Fleischmann, E. D. *J. chem. Phys.* **94**, 3730–3734 (1991).
20. Bug, A. L. R., Wilson, A. & Voth, G. A. *J. phys. Chem.* **96**, 7864–7869 (1992).
21. Cioslowski, J. & Nanayakkara, A. *J. chem. Phys.* **96**, 8354–8362 (1992).
22. Pang, L. & Brisse, F. *J. phys. Chem.* **97**, 8562–8563 (1993).
23. Breton, J., Gonzales-Platas, J. & Girardet, C. *J. chem. Phys.* **99**, 4036–4040 (1993).
24. Isaacs, L., Wehrsig, A. & Diederich, F. *Helv. chim. Acta* **76**, 1231–1250 (1993).
25. Prato, M., Suzuki, T., Wudl, F., Lucchini, V. & Maggini, M. *J. Am. chem. Soc.* **115**, 7876–7877 (1993).
26. Tanner, M. E., Knobler, C. B. & Cram, D. J. *J. org. Chem.* **57**, 40–46 (1992).
27. Bausch, J. W. et al. *J. Am. chem. Soc.* **113**, 3205–3206 (1991).
28. Haddon, R. C. *Science* **261**, 1545–1550 (1993).
29. Diederich, F. & Whetten, R. L. *Acc. Chem. Res.* **25**, 119–126 (1992).
30. Kroto, H. W., Heath, J. R., O'Brien, S. C., Curl, R. F. & Smalley, R. E. *Nature* **318**, 162–163 (1985).

ACKNOWLEDGEMENTS. We thank R. D. Johnson, G. K. S. Prakash, D. J. Cram, R. C. Haddon, and J. Cioslowski for providing information, and the US National Science Foundation for financial support.

Double-walled polymer microspheres for controlled drug release

Kathleen J. Pekarek, Jules S. Jacob & Edith Mathiowitz*

Division of Biology and Medicine, Brown University, Providence, Rhode Island 02912, USA

ONE approach to the controlled release of drugs involves incorporation of the drug molecules into the matrix of microscopic polymer spheres or capsules^{1–10}. Existing methods for preparing such micro-particles do not, however, always guarantee a constant release rate, for example because drug molecules may be trapped preferentially at the surface, because they have to diffuse through an increasing thickness of polymer when the particles are non-eroding or because the surface area changes for eroding particles. In other situations pulsed release may be required—an application to which simple polymer microspheres do not readily lend themselves. Multi-walled microspheres might solve some of these problems. Here we describe a one-step process for preparing double-walled polymer microspheres with diameters ranging from about 20 to 1,000 micrometres. Our technique¹¹ involves the phase separation of a polymer mixture owing to solvent evaporation: with an appropriate choice of interfacial tensions and evaporation rate, a spherical droplet of one polymer becomes coated with a highly uniform layer of the other. This process, which might be adapted to yield multi-walled microspheres, should make possible the engineering of highly specific drug-release properties.

Microspheres containing several shells of different polymers can be produced by pan coating¹⁵, but this does not guarantee uniformly thick layers, which can have disastrous consequences for controlled release. In addition, the process adds a manufacturing step, with attendant quality-control problems and a decrease in yield. Coating processes involving fluidized beds have produced microspheres with a uniform coating¹⁶, but these techniques cannot easily be used to generate particles smaller than about 100 micrometres.

Our approach relies on the spreading equilibria between two fluids suspended as emulsified droplets in a solvent. Harkin's equation¹² describes the tendency of a liquid to spread spontaneously across a solid or liquid surface in terms of the surface and interfacial tensions of the components. This equation can be modified to represent a system in which two dissimilar phases are dispersed within a third by substituting the appropriate interfacial tensions for the surface tensions, leading to¹³

$$\lambda_{AB} = \gamma_{BS} - \gamma_{AS} - \gamma_{AB} \quad (1)$$

* To whom correspondence should be addressed.

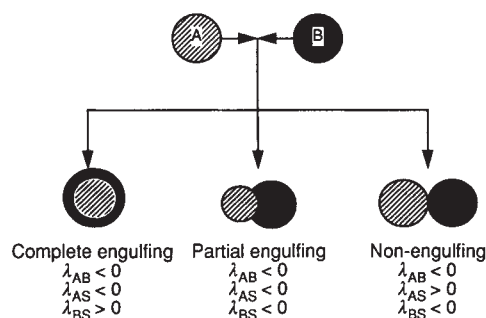


FIG. 1 Three possible configurations for the two-polymer system: complete engulfing, partial engulfing and non-engulfing. The values of the three spreading coefficients (λ_{AB} , λ_{AS} and λ_{BS}) for the system determine which configuration is most thermodynamically stable.

where γ_{AS} and γ_{BS} are the interfacial tensions of the solvent and polymer A or B, respectively, and γ_{AB} is the interfacial tension between polymers A and B. For positive values of the spreading coefficient λ_{AB} , polymer A will spontaneously spread on polymer B. The actual configuration of the two dissimilar phases (A and B) will depend on the values of the three possible spreading coefficients, and can be described as complete engulfing, partial engulfing and non-engulfing (Fig. 1).

Combination of the emulsion theory described above with the phenomenon of phase separation provides a novel approach for preparing double-walled polymer microspheres. By introducing a two-polymer solution into a continuous phase, a stable emulsion is created in which phase separation occurs within each drop. The phase separation can be engineered such that one polymer engulfs the others, forming double-walled microspheres in a single step. Each of the configurations shown in Fig. 1 can be seen in Fig. 2, obtained by scanning electron microscopy (SEM). Cross-sectioned spheres were examined by SEM to determine which polymer dominates a layer by observing differences in surface morphology. SEM also shows other morphological features related to the crystallinity of the polymers such as spherulites. Figure 2a shows a cross-section of a double-walled sphere of poly(1,3-bis(*p*-carboxyphenoxypropane)-co-sebacic anhydride)20:80 and poly(lactic acid) exhibiting phase separation and complete engulfment. Figure 2b shows an SEM image of a partially engulfed or 'hybrid' structure of poly(lactic acid) and poly(ortho ester). Figure 2c shows a non-engulfing, chaotic structure, formed from poly(lactic acid) and polystyrene when precipitation occurred after phase separation but before engulfment.

The microspheres were formed by first preparing separate solutions (10–20% w/v) of the two polymers in a volatile organic solvent such as methylene chloride or chloroform. The drug or protein was added to the appropriate polymer solution and dispersed evenly. Then the two polymer solutions were mixed and dripped into a stirred bath of 0.5% (w/v) poly(vinyl alcohol) in water. As the solvent evaporated, the polymer solution concentration increased to the point where they were no longer mutually soluble and began to separate. If allowed enough time before hardening, the two phases configured themselves according to equation (1) (ref. 13). There was only one thermodynamic equilibrium configuration of polymer A spreading on polymer B, but the extent of spreading was determined by the rate of polymer precipitation, as controlled by the rate of solvent evaporation. Therefore, two routes to double-walled microspheres were possible, and achieved by balancing the thermodynamic and kinetic factors. In the first, polymers for which equation (1) predicts complete engulfment were given sufficient time during the evaporation process for the polymers to reach thermodynamic equilibrium. In this case, complete engulfment was the

most stable thermodynamically, and the kinetic factors were minimized by slowing the rate of polymer precipitation. In the second route, the completely engulfed configuration was a transient intermediate in the progression from a non-phase-separated system toward equilibrium as partially- or non-engulfed, and the precipitation rate was manipulated such that the polymers were

trapped in this configuration, thus allowing the formation of double-walled microspheres when not predicted by equation (1).

Once the microspheres had hardened, they were filtered, washed with water, freeze-dried and stored. They were characterized by optical microscopy (with or without cross-polarized light SEM, transmission electron microscopy (TEM), Fourier-transform infrared microscopy (FTIR) and differential scanning calorimetry (DSC). Optical microscopy proved to be a useful tool to study the microspheres as they were hardening. Cross-polarized light allowed the different polymers to be identified as they crystallized, and fluorescently labelled proteins aided in determining in which layer the protein has been trapped. Once the microspheres became hard, however, the light could not penetrate the spheres and they had to be cross-sectioned for further characterization.

To study further the internal structure of the microspheres, we prepared lyophilized samples of double-walled microspheres for electron and optical microscopy. Semi-thin sections (0.3–0.6 μm) were cut and stained with toluidine blue. Examination of these semi-thin sections with optical microscopy provided important information about the extent of phase-separation. Figure 3a shows semi-thin sections of polyanhydride/poly(lactic acid) microspheres. In a typical experiment, nearly 80% of the

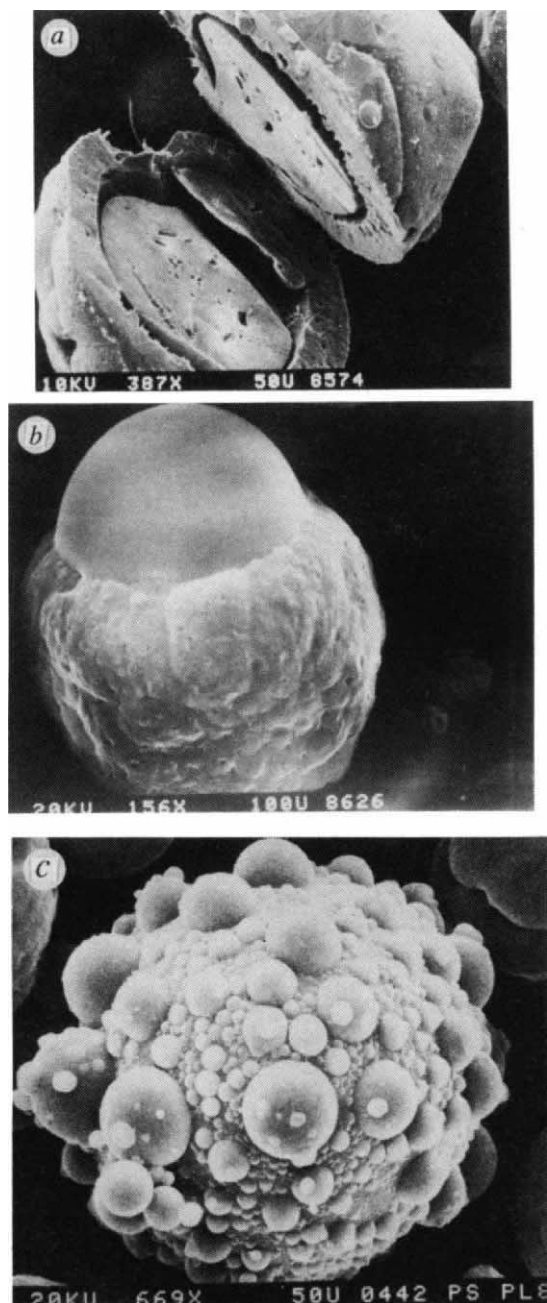


FIG. 2 Scanning electron micrographs of microspheres showing the three configurations. *a*, Cross-section of a double-walled sphere of poly(1,3-bis(*p*-carboxyphenoxypropane)-co-sebacic anhydride):20:80 and poly(lactic acid) showing complete engulfment. *b*, Partially engulfing, or 'hybrid' sphere of poly(lactic acid) and poly(ortho ester). *c*, Non-engulfing structure of poly(lactic acid) and polystyrene. *a*, $\times 236$; *b*, $\times 95$; *c*, $\times 462$.

METHODS. Samples were freeze-dried, mounted on metal stubs and cross-sectioned with a razor blade if necessary. The samples were then sputter-coated with 50–100 Å layer of gold–palladium (Polaron Instrument E5100) and viewed using a Hitachi S-2700 SEM. The poly(ortho ester) was prepared from 3,9-bis(ethylidene-2,4,8,10-tetraoxaspiro-(5,5)undecane), trans-cyclohexane dimethanol and 1,6-hexanediol.¹⁴

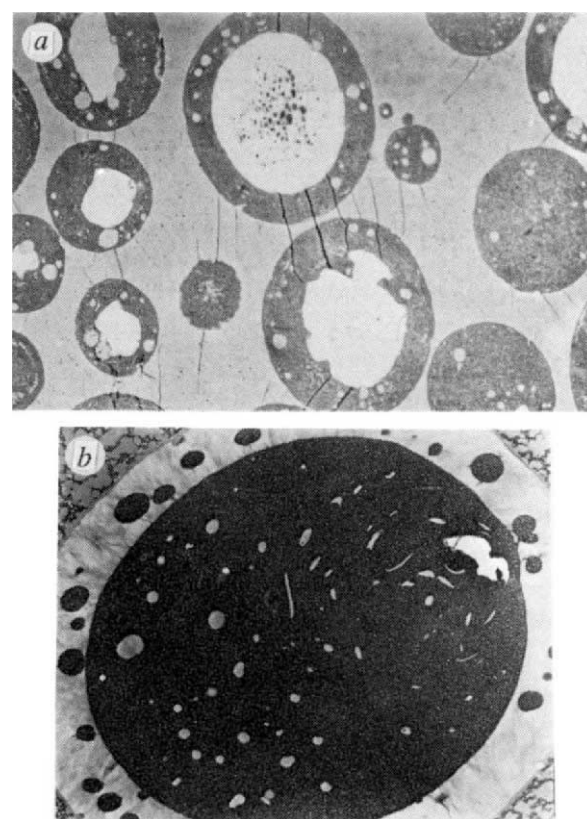


FIG. 3 *a*, Semi-thin section of polyanhydride/poly(lactic acid) microspheres viewed by optical microscopy and *b*, transmission electron micrograph of thin sections of polystyrene/poly(lactic acid) microspheres, showing outer layer and inner core with inclusions of non-phase-separated polymers. *a*, $\times 111$; *b*, $\times 1,950$.

METHODS. Samples were infiltrated with LR white embedding medium for up to 48 h and cured in an oven at 55 °C for 24 h. Semi-thin sections (0.3–0.6 μm) were cut using a glass knife on a Reichert–Jung ultramicrotome and stained using toluidine blue. Thin sections (70–80 nm) were cut using a diamond knife, collected on 200-mesh hexagonal grids, stained in a solution of 0.8% potassium permanganate in 0.1 M potassium phosphate (pH 6.8) and examined on a Phillips EM 410 electron microscope at an accelerating voltage of 100 kV.

beads would phase-separate into double-walled microspheres. Bead size appeared to have little effect on whether or not phase-separation would occur. When examined at higher magnification in the light microscope, it became apparent that in each layer, there existed small domains where phase separation had not been complete. For closer examination of these non-phase-separated domains, thin sections (70–80 nm) of the same sample were also cut, stained and examined by TEM. A potassium permanganate stain provided excellent image contrast, allowing us to visualize the different zones of the microspheres. Figure 3b shows a TEM micrograph of thin sections of polystyrene/poly(lactic acid) microspheres. The outer wall of the microspheres was pale, and contained semi-crystalline or spherulitic domains. The inner-core phase stained intensely and showed a fibrillar matrix. Circular inclusions in the outer wall apparently contained material resembling the fibrillar matrix of the inner core. Similarly, inclusions in the inner core contained material that failed to separate into the outer layer.

The microspheres were also cross-sectioned with a razor blade in order to study the composition of each layer by FT-IR microscopy. By studying the polymer phases as they appeared in the microspheres, the compositions could be compared to the compositions of the layers formed when the solution was left overnight to completely phase-separate. This comparison showed the extent of phase separation that is obtained during the solvent evaporation process and suggested whether the process should be slowed down by reducing the temperature of the non-solvent bath or the volatility of the solvent in order to allow for more complete phase separation.

Finally, the thermal properties of the double-walled microspheres were studied using DSC. For a completely phase-separated system, the DSC thermogram showed the glass transition

temperature and/or melting points of the original polymers. For a system where the polymers were miscible in each other, these temperatures would be shifted, and even merged if the polymers were completely miscible. DSC thermograms of double-walled microspheres showed the two melting points of the original polymers unchanged and distinct, implying that the polymers were immiscible.

Double-layered microspheres can be used to improve present drug delivery technology in a number of ways, including reducing the 'burst effect', fabricating biodegradable and non-degradable spherical devices with zero-order release kinetics and for pulsatile delivery. This new technology for microsphere preparation may have applications in the field of drug or protein delivery. □

Received 25 August; accepted 23 November 1993.

1. Vrancken, M. N. US Patent No. 3523906.
2. Morishita, M. US Patent No. 3960757.
3. Mathiowitz, E., Kline, D. & Langer, R. J. *Scanning Microsc.* **4**, 329–340 (1990).
4. Mathiowitz, E., Saltzman, W. M., Domb, A., Dor, Ph. & Langer, R. J. *appl. Polym. Sci.* **35**, 755–774 (1988).
5. Mathiowitz, E., Dor, Ph., Amato, C. & Langer, R. *Polymer* **31**, 547–555 (1990).
6. Mathiowitz, E. et al. *J. appl. Polym. Sci.* **45**, 125–134 (1992).
7. Madan, P. L. *Drug Dev. Ind. Pharm.* **4**, 95–95–116 (1978).
8. Madan, P. L., Luzzi, L. A. & Price, J. C. *J. phar. Sci.* **61**, 1586–1588 (1972).
9. Heistand, E. N., Wagner, J. G. & Knoechel, E. L. US Patent No. 24899 (1960).
10. Green, B. K. & Schleicher, L. to The National Cash Register Company. US Patent No. 2800457; US Patent No. 2800458 (1963).
11. Mathiowitz, E. & Langer, R. US Patent No. 4861627 (1989).
12. Harkin, W. D. *The Physical Chemistry of Surface Films* (Reinhold, New York, 1952).
13. Torza, S. & Mason, G. J. *Colloid Interface Sci.* **33**, 67–83 (1970).
14. Heller, J., Fritzing, B. K., Ng, S. Y. & Penhale, D. W. J. *Controlled Release* **1**, 87–95 (1985).
15. Parrott, E. L. *Pharmaceutical Technology* 86–88 (Burgess, Minneapolis, Minnesota, 1970).
16. Parrott, E. L. *Pharmaceutical Technology* 90–91 (Burgess, Minneapolis, Minnesota, 1970).

ACKNOWLEDGEMENTS. We thank NOVA Pharmaceutical for providing the polyanhydrides and J. Heller for donating the poly(ortho ester). This work was supported by the Whitaker Foundation.

Effect of deep-sea sedimentary calcite preservation on atmospheric CO₂ concentration

D. Archer* & E. Maier-Reimer†

*† Lamont-Doherty Earth Observatory of Columbia University, Palisades, New York 10964, USA

† Max-Planck-Institut für Meteorologie, Bundesstrasse 7, D-2000 Hamburg 54, Germany

DURING the last glaciation, the atmospheric carbon dioxide concentration was about 30% less than the Holocene pre-industrial value¹. Although this change is thought to originate in oceanic processes², the mechanism is still unclear. On timescales of thousands of years, the pH of the ocean (and hence the atmospheric CO₂ concentration) is determined by a steady-state balance between the supply rate of calcium carbonate to the ocean from terrestrial weathering, and the alteration and removal of carbonate by burial in sediments^{2–4}. Degradation of organic carbon in sediments promotes the dissolution of calcium carbonate in sedimentary pore water^{5,6}, so that a change in the relative rates at which organic carbon and calcium carbonate are deposited on the sea floor should drive a compensating change in ocean pH. Here we use a model that combines ocean circulation, carbon cycling and other sedimentary processes to explore the relationship between deep-sea-sediment chemistry and atmospheric CO₂ concentration. When we include organic-carbon-driven dissolution in our model, a 40% decrease in the calcite deposition rate is enough to decrease the atmospheric CO₂ concentration to the glacial value.

The solubility of the calcite (CaCO₃) in sea water increases with pressure, so that the ocean is typically supersaturated at shallow and intermediate depths and undersaturated in the deepest waters. Only a fraction of the global calcite production is buried, and this proportion depends on the area of the sea floor that is shallower than the depth of calcite saturation. Any imbalance in the sources (terrestrial weathering and alteration) and sinks (deep sea and shallow water deposition) for CaCO₃ will change the ocean dissolved carbonate ion concentration ([CO₃²⁻]) in the direction of restoring throughput balance. Through this mechanism, 'CaCO₃ compensation'^{2–4}, the steady-state deep-sea calcite burial rate is determined by the rate of weathering less shallow water deposition, which we refer to as the 'deep-sea carbonate influx'. The pH equilibrium reaction CO₂ + CO₃²⁻ + H₂O ↔ 2HCO₃⁻ maintains an inverse relationship between [CO₃²⁻] and pCO₂.

We simulate the coupled ocean-sediment carbon cycle using the global ocean circulation-carbon cycle model of Maier-Reimer^{7,8}, to which we have coupled the sediment calcite dissolution model of Archer⁶. Without the detailed sedimentary component the ocean model is unable to drive the atmospheric pCO₂ to lower glacial values without violating palaeoceanographic data⁷. Here we predict the ocean alkalinity ([HCO₃⁻] + 2[CO₃²⁻] + [B(OH)₄⁻]) and total CO₂ ([CO₂] + [HCO₃⁻] + [CO₃²⁻]), rather than specifying present-day values, based on the constraint that the deep-sea calcite burial rate equals the deep-sea carbonate influx (which is imposed as a source of alkalinity and total CO₂ to the surface ocean, in the ratio of 2:1). The deep-sea alkalinity influx for the 'standard case' (intended to mimic present-day conditions) is 1.2 Gt-CaCO₃ yr⁻¹ (refs 9, 10). The model distributions of alkalinity and total CO₂, and therefore the calcite saturation state of deep water, compare well with oceanic observations (Fig. 1a and refs 7 and 8). Although the glacial distributions of nutrients, δ¹³C

† Present address: Department of Geophysical Sciences, 5734 South Ellis Avenue, University of Chicago, Chicago, Illinois 60637, USA.

and sedimentary calcite were influenced by changes in circulation, we use the present-day ocean circulation field throughout.

The calcite diagenesis model⁶ calculates the calcite dissolution rate from the steady-state concentration profiles of the carbonate species CO_2 , HCO_3^- and CO_3^{2-} in the diffusive sediment pore water. Within the constraints of *in situ* microelectrode data¹¹, sediment trap data, pore water chemical data and calcite accumulation rates, the model predicts the observed patterns of calcite preservation in the ocean⁶ (Fig. 1b). The steady-state sedimentary calcite distribution in the 'standard case' model is compared with observations in Fig. 2a and b. The model calcite distribution is smoother than observations because of the smoothness of the model topography.

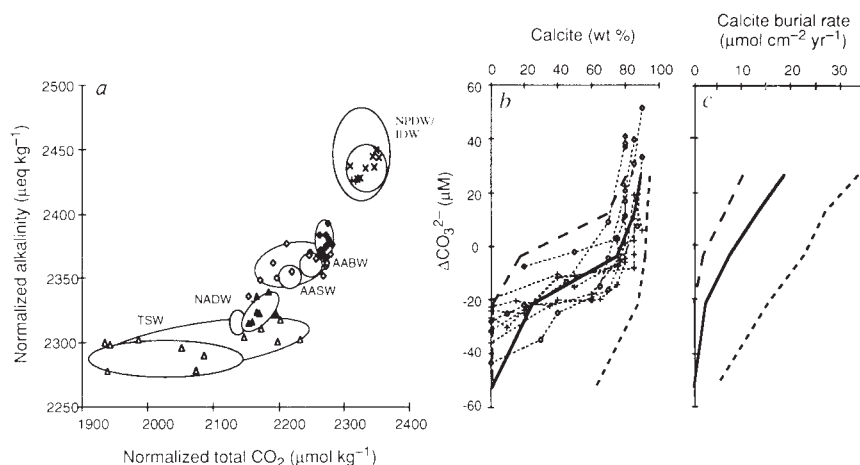
One proposed explanation for lower glacial $p\text{CO}_2$ is called the 'coral reef hypothesis'^{12,13}. In the present-day ocean, up to half of the total oceanic CaCO_3 removal may be by growth of coral reef complexes¹³. The rate of coral growth might be tied to fluctuations in sea level relative to the level of the continental shelves, with low coral growth rates during glacial times when the shelves are exposed and surface temperatures are cooler. A shift of the 'locus of CaCO_3 deposition' to the deep sea would require a higher ocean $[\text{CO}_3^{2-}]$, and hence a lower $p\text{CO}_2$ (refs 12, 13).

FIG. 1 a, Comparison of the ocean models with oceanographic data, showing the distribution of alkalinity and total CO_2 in the ocean water masses, normalized to a salinity of 35 parts per thousand. Symbols are selected data from GEOSECS³⁵, outlined by unshaded ovals. Shaded ovals are from the 'standard case' model runs. The model captures the general increase in dissolved carbon alkalinity with age since contact with the atmosphere, as tropical surface water is subducted into the deep sea as North Atlantic deep water, into Antarctic surface water and Antarctic deep water, to Indian and Pacific deep waters. b and c, Comparison of the sedimentary calcite preservation model⁶ with oceanographic data. Vertical axis is ΔCO_3^{2-} , the saturation state of the water column with respect to calcite; because of the pressure dependence of calcite solubility²⁰, $\mu\text{M } \Delta\text{CO}_3^{2-}$ corresponds to approximately 1 km depth. Rain rates of organic carbon and calcite are from the circulation model; see above (solid line; standard case: long dashes; high organic rain: short dashes; no organic carbon rain). b, Calcite (dry wt%), from ref. 6. Hollow diamonds indicate summarized data from the Atlantic Ocean, crosses indicate data from the Pacific Ocean. Scatter in the data may reflect variability in organic and calcite rain rates. Calcite dissolution kinetics follow ref. 30, using a rate constant of 1 d^{-1} and a reaction order of 4.5. Organic carbon respires with O_2 according to the first-order rate constant $2 \times 10^{-9} \text{ s}^{-1}$. Anoxic diagenesis is neglected. Within the sediment mixed layer (taken to be 10 cm deep), calcite is treated as well-mixed; the solid organic carbon concentration profile is calculated using a mixing coefficient of $150 \text{ cm}^2 \text{ kyr}^{-1}$. c, Burial rate of calcite. By observing that the standard- and the high-organic carbon cases are similar to each other but offset by $20 \mu\text{M}$, we can anticipate that the high-organic rain scenario will require an increase in $[\text{CO}_3^{2-}]$ of at least $20 \mu\text{M}$ to maintain constant CaCO_3 burial rate. Through pH equilibrium chemistry, an increase of $[\text{CO}_3^{2-}]$ of this magnitude should lead to a decrease in $p\text{CO}_2$ of $\sim 20\%$ ($\sim 60 \mu\text{atm}$), in a homogeneous ocean. Actual global carbon cycle model response to the increase in organic rain was a $[\text{CO}_3^{2-}]$ increase of $29 \mu\text{M}$, and a $p\text{CO}_2$ decrease of $55 \mu\text{atm}$.

METHODS. For a, model resolution is 3.3° and the time step is one year. The rate of organic production at the surface is determined by the rate of supply of phosphate from deep water. The production of calcite is taken to be 25% of the organic carbon production on a molar basis, except in very cold waters, where calcite production is diminished²⁸. The global organic carbon export production in the model is 7.3 Gt-C yr^{-1} (1 Gt is 10^{15} g). Local new production rates range from the $5 \text{ mol-C m}^{-2} \text{ yr}^{-1}$ in the equatorial Pacific to a minimum of

The steady-state $p\text{CO}_2$ predicted by our model appears to respond linearly to variations in the deep-sea carbonate influx (Fig. 3a and b). In order for our model to achieve the magnitude of the observed glacial/interglacial $p\text{CO}_2$ changes, the glacial deep-sea calcite burial rate has to be 2–3 times higher than today's value. Under these conditions, high-calcite sediments dominate the sea floor (Fig. 2c). This prediction is clearly at odds with the sedimentary record^{14,15}. Although a better test of the coral-reef hypothesis would be a time-dependent calculation¹³, our steady-state model is justified by the observation that the atmospheric $p\text{CO}_2$ recorded by the Vostok ice core was depressed to the $180\text{--}220 \mu\text{atm}$ range for 50 kyr (ref. 16). This time period is long enough for sediments to have approached their steady-state condition. Our results support the conclusion that cycles in the global growth-rate of corals would have a significant impact on $p\text{CO}_2$; however, we show that the coral-reef mechanism by itself cannot account for the entire glacial/interglacial $p\text{CO}_2$ signal.

We propose a new explanation for lower glacial $p\text{CO}_2$ that relies in part on the effect of sedimentary organic carbon degradation on calcite dissolution. Diagenetic models for calcite dissolution predict that a significant fraction of the calcite rain to the



$\sim 0.3 \text{ mol-C m}^{-2} \text{ yr}^{-1}$ in the northeast Pacific. The rain rate of organic carbon at depth was determined based on the production rate, in the form suggested in ref. 29.

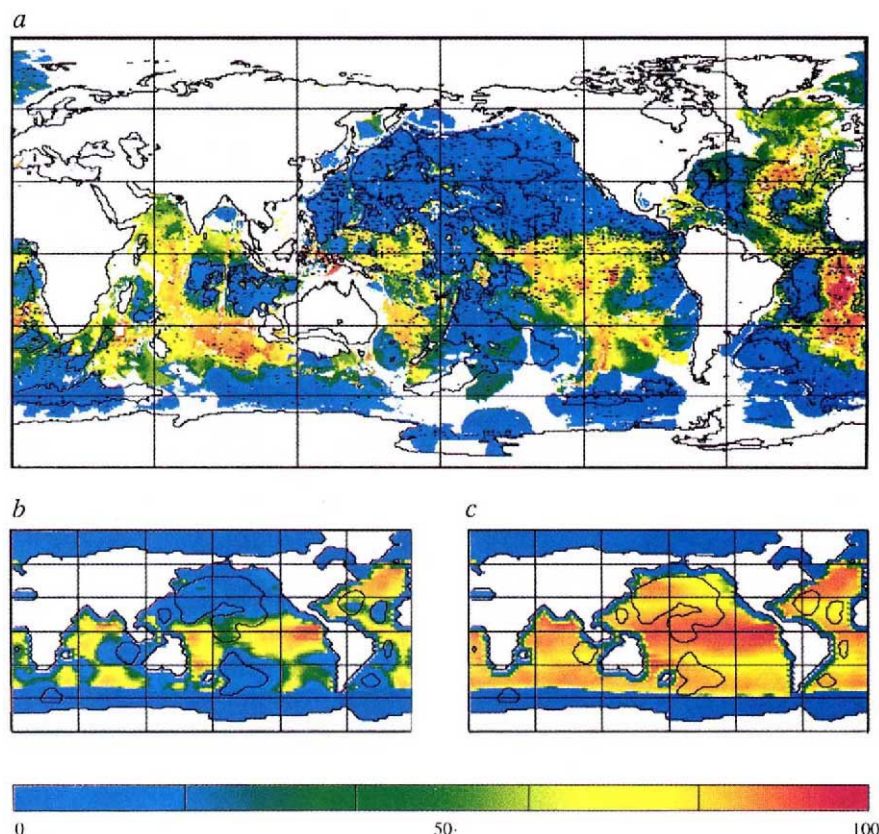
$$F_{oc} = F_{oc}(25) (z/100)^{-n} \quad (1)$$

where F_{oc} is the sinking flux of organic carbon at any depth, $F_{oc}(25)$ is the flux from the surface box and z is the depth in metres. The value for n is taken to be 0.7 for the standard case. The organic rain to sediments ranged from 0.02 to $0.40 \text{ mol-C m}^{-2} \text{ yr}^{-1}$, with a global average value of $0.14 \text{ mol-C m}^{-2} \text{ yr}^{-1}$. An analogous expression for the calcite rain rate takes the form:

$$F_{ca} = XF_{ca}(25) + (1-X) F_{ca}(25) \exp(-z/2000) \quad (2)$$

where $X=0.3$ is the proportion of the calcite production flux ($F_{ca}(25)$) that sinks (with no attenuation) to the sediment, and the rest (0.7) dissolves in the water column with an e-folding depth of 2,000 m. The molar ratio of organic carbon flux to calcite flux (to the sediments) ranged from 0.65 to 0.8 (refs 5, 6). The sedimentary rain rate of terrigenous (refractory) material is taken to be a uniform $1.8 \times 10^{-5} \text{ g cm}^{-2} \text{ yr}^{-1}$ over the entire pelagic sea floor (with values two orders of magnitude higher in locations adjacent to the continents). The $p\text{CO}_2$ of the model atmosphere is determined by gas exchange with the sea surface. In order to sustain productivity in the gyres, far from areas of upwelling, the model requires sea surface nutrient concentrations that are higher than those observed (consistent with ref. 28; this tends to artificially offset the model atmospheric $p\text{CO}_2$ by approximately $20 \mu\text{atm}$).

FIG. 2 Distribution of sedimentary calcite (dry wt%; see colour bar at bottom of figure); comparison of model results with oceanographic data. The contour (in a, b and c) indicates the 5 km isobath. a, An interpolation from oceanographic data^{31–33}. Black dots indicate data locations. The colour field was generated using a 0.5° bathymetry and depth interpolating to the nearest data; the white fields are too far from data, in depth or horizontally, for extrapolation. b, Distribution of calcite predicted by the model, for the standard case. The model bathymetry is smoother than the bathymetry used to construct a, which results in a smoother distribution of calcite. c, Model distribution of calcite that results from an increased deep-sea carbonate influx (260% higher than the standard case). As sediments preserved from the last glacial maximum are not uniformly high in calcite, we conclude a lower glacial rates of shallow-water deposition is not solely responsible for low glacial atmospheric $p\text{CO}_2$ values.

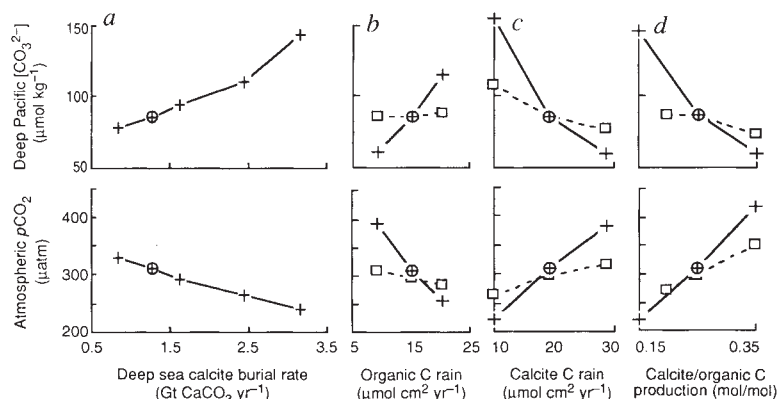


sediments dissolves in response to the addition of CO_2 to the pore water by oxic organic carbon degradation^{5,6} (a process we will refer to as 'respiratory calcite dissolution', see Fig. 1b and c). If sedimentary calcite dissolution is represented as the sum of respiratory dissolution and dissolution driven by the overlying water $[\text{CO}_3^{2-}]$, a global increase in one of these quantities will

require a decrease in the other to maintain a steady state. Therefore an increase in the organic carbon/calcite ratio of the material reaching the sediments will drive an increase in $[\text{CO}_3^{2-}]$, and a corresponding decrease in $p\text{CO}_2$, through the mechanism of CaCO_3 compensation.

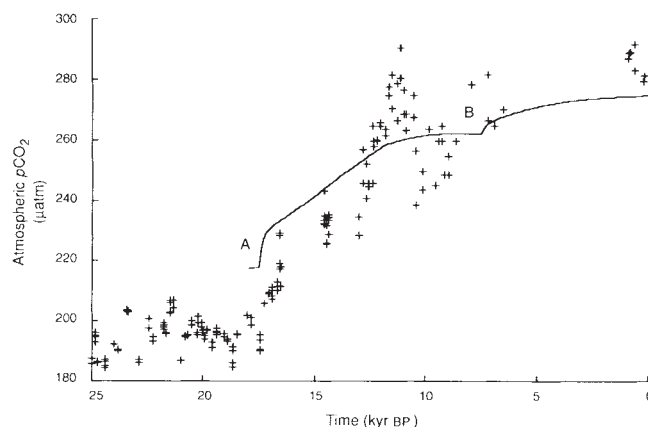
We test this hypothesis by altering the fraction of organic

FIG. 3 Steady-state response of the coupled model to changes in the forcings for calcite burial: a comparison of results including and neglecting respiratory calcite dissolution. Upper plots are model concentration of $[\text{CO}_3^{2-}]$ in the deep Pacific (Equator and dateline, 5,000 m depth). Lower plots are atmospheric $p\text{CO}_2$. Solid lines with crosses are model results including the effect of organic carbon respiration in the sediments, with the 'standard case' circled; dashed lines with hollow squares are corresponding results neglecting respiratory calcite dissolution, and re-tuned to achieve a realistic lysocline and deep-sea burial rate (parameter X in equation (2), Fig. 1, was lowered from 0.3 to 0.07, and the dissolution rate constant was increased from 1 to 100 d^{-1}). a, Model response to the deep-sea carbonate influx. The distribution of calcite for the highest value used ($3.1 \text{ GT-CaCO}_3 \text{ yr}^{-1}$) is shown in Fig. 2c. b, Model response to the organic-carbon supply rate to the sea floor (average flux at 4,500 m depth). The fraction of organic carbon production that reaches the sea floor was adjusted by changing the value of n (equation (1), Fig. 1) from the standard value of 0.7 to 0.6 (high organic rain) and 0.85. Altering the dynamics of nutrient recycling changed the organic production rate by $\pm 9\%$; however the global calcite-production rate was held to the standard value. Sensitivity of model atmospheric $p\text{CO}_2$ to organic-carbon sedimentary flux depends on respiratory calcite dissolution. c, Model response to the average calcite supply rates to the sea floor. The calcite sinking flux was adjusted by changing the value of the parameter X through the range (0.1, 0.3, 0.5) for the runs including organic carbon respiration, and (0.03, 0.07, 0.12) for the runs neglecting organic carbon sediment chemistry, where the middle values are the 'standard' values in each



case. The effect of calcite supply is in the opposite direction to that of organic carbon: an increase of production drives an increase in $p\text{CO}_2$. This effect is stronger when respiratory calcite dissolution is included in the model. d, Model response to variation in the ratio of calcite to organic-carbon production in surface waters, with 0.25 being the standard (present-day) value. When respiratory calcite dissolution is neglected and a ratio of 0.15 is used, the sinking flux of calcite to the deep sea is insufficient to balance weathering and no steady state exists for this model. When respiratory calcite dissolution is included in the simulation, a 40% decrease in calcite production is sufficient to drive atmospheric $p\text{CO}_2$ close to glacial values.

FIG. 4 Time-dependent evolution of model $p\text{CO}_2$ response to a 'glacial termination' scenario, compared with data from the Byrd ice core¹ plotted on the new timescale of Sowers (T. Sowers, personal communication). Initial conditions were the steady-state response to low calcite/organic carbon production rate ratio of 0.15 (Fig. 2e, f) and a reconstruction of the glacial calcite distribution (based on the generalization that calcite was preserved 350 m deeper in the glacial Pacific Ocean, 600 m shallower in the Atlantic Ocean, and the same as present day in the Indian Ocean⁴). At 17.5 kyr, the production ratio was restored to the 'standard' value of 0.25, and a 10 kyr draw-down of CO_2 by forest regrowth was initiated. The initial, fast rise in $p\text{CO}_2$ at 17.5–17.3 kyr BP (point A) is caused by a change in surface water chemistry; note that if surface ocean chemical rearrangement were entirely responsible for the low glacial $p\text{CO}_2$, the transitions would presumably be much faster than observed in the data. The sudden (artificial) end of the forest regrowth sink for CO_2 at 7.5 kyr BP (point B) can be seen in the kink in $p\text{CO}_2$ at that point.



carbon and calcite production that reaches the sea floor (Fig. 3b and c). As only a small fraction of production reaches the ocean bottom, altering this fraction has only a minor impact on the tracer source/sink functions in the water column. In order to isolate the effect of respiratory calcite dissolution, we ran model simulations both with and without this process. When respiratory calcite dissolution is neglected, we find that $p\text{CO}_2$ is sensitive to calcite rain to the sea floor, and insensitive to organic carbon rain, consistent with previously studies¹⁷. In contrast, when respiratory calcite dissolution is included in the simulation, $p\text{CO}_2$ becomes much more sensitive to organic carbon and calcite rain rates (Fig. 3b–d, solid lines). We also simulate an ecological shift from calcite to siliceous organisms by varying the production ratio of calcitic to organic carbon. For example, a decrease in calcite production by 40% globally drives the $p\text{CO}_2$ of the atmosphere near glacial values.

The required shifts in organic carbon and calcite dissolution are qualitatively supported by observations from glacial sediments. An increase in organic carbon production is suggested by an increased benthic-planktonic gradient in $\delta^{13}\text{C}$ (ref. 18), increased sedimentary organic carbon concentration¹⁹ and burial rates²⁰, and a change in foraminiferal assemblages²¹. In general, calcitic organisms are replaced by siliceous diatoms in regions of high productivity and lower temperatures²². Both conditions are thought to be generally more prevalent under the glacial climate. In high latitudes, the front between silica and calcite production moved towards the Equator²³ during the glacial. Thus we see a general increase in organic production, and a potential shift from calcite toward siliceous production during the last glacial. We conclude that, in contrast to the coral-reef hypothesis, the requirements of a 'change in production' hypothesis seem to fit more comfortably within the constraints of the available data.

The model predicts that the alkalinity and pH of the whole ocean were higher during glacial time than at present. One test of this hypothesis will be palaeo-pH estimates derived from isotopic signature of boron^{24,34} in sedimentary calcite (A. Sanyal and G. Hemming, personal communication). Another test of the hypothesis can be made by comparing model time-dependent behaviour to atmospheric $p\text{CO}_2$ recorded in ice cores. The ~10-kyr timescale of data from Byrd¹ is clearly longer than would be expected from a glacial/interglacial change in surface ocean chemistry^{25,26} or ocean circulation¹⁷. The model experiment in Fig. 4 started at 18 kyr before present, (BP) with a low value for calcite production relative to organic carbon, and simulated the removal of carbon from the ocean by forest regrowth (as inferred from the $\delta^{13}\text{C}$ values of benthic foraminifera²⁷) by removing 0.055 Gt-C yr⁻¹ from the atmosphere, from 17.5 to 7.5 kyr BP, for a total of 550 Gt-C. The model captures the long-term behaviour of the observed CO_2 record, but misses short-term excursions

(for example at 12 kyr BP) that are probably associated with fast changes in ocean circulation, sea surface temperature or terrestrial biomass, that are not simulated by the model. The model also neglects changes in terrestrial weathering intensity and coral growth that may be significant in the global alkalinity budget. We conclude, however, that the similarity between the model $p\text{CO}_2$ and data from Byrd supports the possibility that lower glacial $p\text{CO}_2$ values were caused by a higher glacial whole-ocean alkalinity. Because the 'coral reef' hypothesis (a higher deep-sea calcite burial rate during glacials) appears by itself to be inadequate as an explanation for high glacial alkalinity, we look instead to a change in the relative production rates of calcite and organic carbon. □

Received 8 February; accepted 23 November 1993.

1. Neftel, A., Oeschger, H., Staffelbach, T. & Stauffer, B. *Nature* **331**, 609–611 (1988).
2. Broecker, W. S. & Peng, T. H. *Global Biogeochem. Cycles* **1**, 15–29 (1987).
3. Boyle, E. A. *Nature* **331**, 55–56 (1988).
4. Emerson, S. R. & Archer, D. E. *Paleoceanography* **7**, 319–332 (1992).
5. Emerson, S. & Bender, M. L. *J. mar. Res.* **39**, 139–162 (1981).
6. Archer, D. E. *J. geophys. Res.* **96**, 17037–17050 (1991).
7. Heinze, C., Maier-Reimer, E. & Winn, K. *Paleoceanography* **6**, 395–430 (1991).
8. Maier-Reimer, E. *Global Biogeochem. Cycles* **7**, 645–678 (1993).
9. Milliman, J. D. *Marine Carbonates 1–375* (Springer, Heidelberg, 1974).
10. Davies, T. A. & Worsley, T. R. *Paleoenvironmental Implications of Oceanic Carbonate Sedimentation Rates*, 169–179 (Spec. Publ. No. 32, Soc. Econ. Paleontologists and Mineralogists, Tulsa, 1981).
11. Archer, D., Emerson, S. & Reimers, C. *Geochim. cosmochim. Acta* **53**, 2831–2846 (1989).
12. Berger, W. H. *Naturwissenschaften* **69**, 87–88 (1982).
13. Opdyke, B. N. & Walker, J. C. G. *Geology* **20**, 733–736 (1992).
14. Berger, W. H. & Keir, R. S. in *Climate Processes and Climate Sensitivity* (eds Hansen J. E. & Takahashi, T.) 337–351 (Am. geophys. Un., Washington DC, 1984).
15. Peterson, L. C. & Prell, W. L. in *The Carbon Cycle and Atmospheric CO_2 : Natural Variations Archae to Present* (eds Sundquist E. T. & Broecker, W. S.) 251–269 (Am. Geophys. Un., Washington DC, 1985).
16. Barnola, J. M., Raynaud, D., Korotkevich, Y. S. & Lorius, C. *Nature* **329**, 408–414 (1987).
17. Keir, R. S. *Paleoceanography* **5**, 253–277 (1990).
18. Sarnthein, M., Winn, K., Duplessy, J. C. & Fontugne, M. R. *Paleoceanography* **3**, 361–399 (1988).
19. Lyle, M. et al. *Paleoceanography* **3**, 39–59 (1988).
20. Lyle, M. *Nature* **335**, 529–532 (1988).
21. Mix, A. C. *Nature* **337**, 541–544 (1989). (OK?)
22. Lisitzin, A. P. in *The Micropaleontology of Oceans* (eds Funnell, B. M. & Reidel, W. R.) 173–195 (Cambridge Univ. Press, London, 1971).
23. Howard, W. R. & Prell, W. L. *Paleoceanography* **7**, 79–118 (1992).
24. Spivack, A. J., You, C.-F. & Smith, H. J. *Nature* **363**, 149–151 (1993).
25. Broecker, W. S. & Peng, T.-H. *Global Biogeochem. Cycles* **3**, 215–239 (1989).
26. Sarmiento, J. L., Toggweiler, J. R. & Najjar, R. *Phil. Trans. R. Soc. Lond. A* **325**, 3–21 (1988).
27. Leuenberger, M., Seigenthaler, U. & Langway, C. C. *Nature* **357**, 488–490 (1992).
28. Bacastow, R. & Maier-Reimer, E. *Clim. Dynam.* **4**, 95–125 (1990).
29. Martin, J. H., Knauer, G. A., Karl, D. M. & Broenkow, W. M. *Deep-Sea Res.* **34**, 267–285 (1987).
30. Keir, R. S. *Geochim. cosmochim. Acta* **44**, 241–252 (1980).
31. Biscaye, P. E., Kolla, V. & Turekian, K. K. *J. geophys. Res.* **81**, 2595–2603 (1976).
32. Berger, W. H., Adelseck, C. G. & Mayer, L. A. *J. geophys. Res.* **81**, 2617–2672 (1976).
33. Kolla, V., Bé, A. W. H. & Biscaye, P. J. *J. geophys. Res.* **81**, 2605–2616 (1976).
34. Hemming, N. G. & Hanson, G. N. *Geochim. cosmochim. Acta* **56**, 537–544 (1992).
35. Takahashi, T., Broecker, W. S., Bainbridge, A. C. & Weiss, R. F. *Tech. Rep. No. 1, CU-1-80* (Lamont-Doherty Geological Observatory, Palisades, New York, 1980).

ACKNOWLEDGEMENTS. We thank W. Broecker, S. Emerson, L. Burkle, R. Keir, M. Lyle and D. MacAyeal for discussions. This work was supported by the Lamont Doherty Earth Observatory.

Geodetic slip rate for the eastern California shear zone and the recurrence time of Mojave desert earthquakes

Jeanne Sauber*, Wayne Thatcher†, Sean C. Solomon‡ & Michael Lisowski†

* Geodynamics Branch, Laboratory for Terrestrial Physics, NASA Goddard Space Flight Center, Greenbelt, Maryland 20771, USA

† US Geological Survey, Menlo Park, California 94025, USA

‡ Department of Terrestrial Magnetism, Carnegie Institution of Washington, Washington DC 20015, USA

WHERE the San Andreas fault passes along the southwestern margin of the Mojave desert, it exhibits a large change in trend, and the deformation associated with the Pacific/North American plate boundary is distributed broadly over a complex shear zone. The importance of understanding the partitioning of strain across this region, especially to the east of the Mojave segment of the San Andreas in a region known as the eastern California shear zone (ECSZ), was highlighted by the occurrence (on 28 June 1992) of the magnitude 7.3 Landers earthquake in this zone. Here we use geodetic observations in the central Mojave desert to obtain new estimates for the rate and distribution of strain across a segment of the ECSZ, and to determine a coseismic strain drop of ~ 770 μrad for the Landers earthquake. From these results we infer a strain energy recharge time of 3,500–5,000 yr for a Landers-type earthquake and a slip rate of ~ 12 mm yr^{-1} across the faults of the central Mojave. The latter estimate implies that a greater fraction of plate motion than heretofore inferred from geodetic data is accommodated across the ECSZ.

Although most of the relative motion between the Pacific and North American tectonic plates in the southwestern United States occurs by slip on the San Andreas fault, there are numerous other faults that have slipped in the past million years^{1,2}. Very-long-baseline-interferometry (VLBI) measurements made since 1984 (Fig. 1) and conventional laser ranging measurements begun in 1971 (ref. 3) show that 75% of the continuing motion between the Pacific and North American plates is distributed across a zone ~ 100 km wide in southernmost California. The strike of the San Andreas fault is approximately parallel to the slip direction of relative plate motion, N33–41° W (ref. 4), along most parts of the plate boundary. But along a 190-km-long segment between San Geronio Pass and Tejon Pass (referred to here as the Mojave segment) the San Andreas fault locally changes its trend by 20° to 30° (Fig. 1). Across this segment of the plate boundary, the width of the zone accommodating a comparable fraction of plate motion is broader (~ 200 km) and the mode of deformation is more complex (Fig. 1). To the northeast of the Mojave segment of the San Andreas fault are the right-lateral strike-slip faults of the Mojave desert (Fig. 1). Palaeoseismic⁵, geological⁶, and geodetic^{7,8,9} observations suggest the presence of a zone of shear-strain accumulation which extends from the San Geronio Pass to north of the Garlock fault in eastern California, known as the eastern California shear zone (ECSZ). Aftershocks from the Landers earthquake and an adjacent magnitude 6.1 event in April 1992 defined a seismic lineament >80 km long within the ECSZ.

The central Mojave geodetic network spans a segment of the ECSZ from west of the Helendale fault to east of the Ludlow fault and includes the northern segment of the Landers earthquake rupture (Fig. 1). We have determined the rate and spatial distribution of continuing strain for a number of different spatial subnets using a moving-window approach. An earlier analysis of the same data⁷ was adversely affected by a poor

eccentric reduction between observations made in 1935 and those made in 1965 and 1982 at a station just east of the Camp Rock fault, and by three angles with unusually large angle changes (>4 arcsec) involving a station just east of the Pisgah fault. Whereas the earlier analysis⁷ suggested that an abrupt decrease in the rate of shear strain occurs to the east of the Camp Rock fault (Fig. 1), our new results (with the poor data removed) suggest that active strain accumulation extends at least between the Helendale and Pisgah faults, a zone ~ 60 km wide. Between the Pisgah and Ludlow faults we were unable to discern a significant strain rate (Table 1). The average shear-strain rate between the Helendale and Pisgah faults during 1934–82 was $\dot{\gamma} = 0.15 \pm 0.03$ $\mu\text{rad yr}^{-1}$, with the maximum right-lateral shear strain occurring on a plane oriented N39° W $\pm 5^\circ$. Geological observations suggest that a small component of north–south contraction accompanies slip on the northwest-striking faults¹⁰. The geodetic observations from the central Mojave network

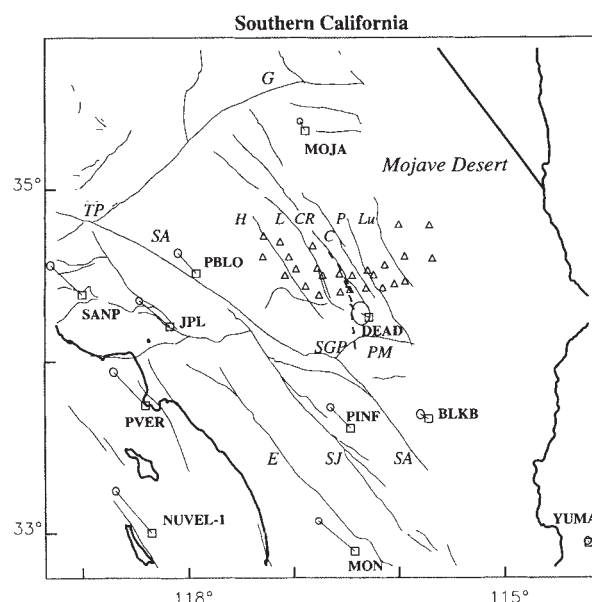


FIG. 1 The location of the stations in the central Mojave geodetic network (triangles), the 28 June 1992 Landers earthquake rupture (dashed line)²⁶, and the vector velocity of western US sites (squares) and associated 95% confidence ellipses estimated from VLBI measurements. We determined site velocities from VLBI data²⁷ taken over the period 1984–91 in a reference frame defined by minimizing the velocity of reference sites (North American plate: Haystack and Westford, Massachusetts; Pietown, New Mexico; Platteville, Colorado; Gilcreek, Alaska. Pacific plate: Kauai, Hawaii; Vandenberg, California. Eurasian plate: Wettzell, Germany; Onsala, Sweden.) with respect to the predicted NUVEL-1 (ref. 4) velocity in a frame in which North America is fixed. The slip direction and rate (49.6 ± 1.3 mm yr^{-1}) of Pacific plate motion with respect to the North American plate predicted by the global plate motion model NUVEL-1 is given in the lower left corner. Faults or fault zones¹: C, Calico; CR, Camp Rock-Emerson; E, Elsinore; G, Garlock; H, Helendale; L, Lenwood; Lu, Ludlow; PM, Pinto Mountain; P, Pisgah; SA, San Andreas; SJ, San Jacinto. Other features: SGP, San Geronio Pass; TP, Tejon Pass. The velocities of some of the VLBI stations, with their 1σ uncertainties are: DEAD, 7.7 ± 3.4 mm yr^{-1} at N62° W $\pm 15^\circ$; MOJA, 9.5 ± 1.1 mm yr^{-1} at N26° W $\pm 3^\circ$; PBLO, 24.4 ± 1.4 mm yr^{-1} at N41° W $\pm 2^\circ$; YUMA, 1.8 ± 1.3 mm yr^{-1} at N34° W $\pm 20^\circ$. The differenced velocity vector between MON and BLKB is 33.5 ± 1.9 mm yr^{-1} at N48° W $\pm 2^\circ$, between MON and YUMA is 39.4 ± 1.7 mm yr^{-1} at N50° W $\pm 1^\circ$, between JPL and PBLO is 11.1 ± 1.8 mm yr^{-1} at N67° W $\pm 5^\circ$, between JPL and MOJA is 26.2 ± 1.6 mm yr^{-1} at N57° W $\pm 2^\circ$, and between PVER and MOJA is 31.3 ± 1.8 mm yr^{-1} at N48° W $\pm 2^\circ$.

TABLE 1 Rate and orientation of maximum right-lateral shear strain, central Mojave network, 1934–82

Region	No. of obs.	$\dot{\gamma}$ ($\mu\text{rad yr}^{-1}$)	ψ (degrees)
Helendale/Lenwood	28	0.14 ± 0.04	-45 ± 12
Camp Rock	12	0.16 ± 0.06	-44 ± 9
Landers rupture	10	0.22 ± 0.09	-42 ± 7
Calico/Pisgah	12	0.15 ± 0.06	-28 ± 15
Pisgah/Ludlow	18	0.06 ± 0.06	unresolved
Helendale/Pisgah	69	0.15 ± 0.03	-39 ± 5
Helendale/Ludlow	92	0.14 ± 0.03	-43 ± 5

The maximum horizontal shear-strain rate is denoted by $\dot{\gamma}$, and ψ (positive clockwise from N) is the azimuth of maximum right-lateral shear strain for subnetworks straddling the named faults. Angle changes over the period 1934–82 were obtained from historical triangulation measurements made between 1934 and 1940 and again in 1965 and trilateration measurements made in 1982. To estimate strain rates, angle changes between successive surveys were fitted to a strain-rate field assumed to be spatially and temporally uniform using a modified version of Frank's method²⁵.

indicate that the axis of maximum compressional strain is orientated N06° W $\pm$ 5°, similar to the observed direction of contraction across local folds (A. Glazner, personal communication). As geological data and recent trenching studies^{11,12} indicate large right-lateral displacements along faults trending north-to-northwest, and right-lateral slip occurred on some of these faults in the Landers event, $\dot{\gamma}$ is most simply interpreted as representing right-lateral shear strain.

The geodetically measured strain rate in the central Mojave desert could be due to one or more of several different mechanisms: (1) earthquakes, (2) fault creep, (3) elastic strain, or (4) anelastic processes such as folding. We had postulated earlier that most of the measured strain between the Helendale and Camp Rock faults was the result of either elastic strain accumulation or fault creep⁷. Although a part of the observed straining across the western part of the central Mojave network could be due to strain accumulation across the San Andreas fault, the magnitude of this effect is uncertain. A dislocation model of the San Andreas fault with a locking depth of 20 km (the maximum depth of seismicity¹³) and a slip rate of 30 mm yr⁻¹ (ref. 14) suggests that elastic strain associated with the San Andreas fault could account for as much as 30% of strain rate in the Helendale/Lenwood subnetwork but <10% of that across the Calico/Pisgah subnetwork. The amount of this San Andreas contribution depends on uncertain modelling assumptions, and so we have made no attempt to correct for it in what follows. If significant, its effect would be to decrease estimates of cumulative slip across central Mojave faults and increase estimates of strain recharge times across the Landers rupture. That the rupture during the Landers earthquake extended across numerous fault segments confirms that the strain rate measured over a broad region in the Mojave desert is released, at least in part, by large Mojave earthquakes. An independent assessment of the seismic hazard in southern California based on geological offsets also led to the suggestion that major earthquakes (with magnitude up to ~6.9) would occur on individual fault segments of the Mojave desert¹⁵.

Within the brittle upper crust, tectonic motion across a plate boundary is accommodated primarily by earthquakes or episodic creep. Below the seismogenic zone, slip has been hypothesized to occur aseismically on a downward extension of individual faults, or as a broad shear zone below the surface trace of several faults¹⁶. A shear zone model with multiple discrete faults underlain by a zone of distributed ductile deformation can be used to represent the rate and distribution of strain in the central Mojave and to estimate an equivalent slip rate across the shear zone. The engineering shear-strain rate $\dot{\gamma}$ at the surface is related to the distance x from the centre of the

shear zone by¹⁶:

$$\dot{\gamma} = \frac{-\dot{b}}{2\pi W} \left(\tan^{-1} \frac{x-W}{D} - \tan^{-1} \frac{x+W}{D} \right)$$

where $\dot{b}$ is the slip rate across the zone of deformation, D is the depth of the shear zone, and W is the half-width of the shear zone. The depth of background seismicity (the maximum focal depth is 15 km, and most events are at <6 km (ref. 7) and Landers aftershocks¹⁷ and the depth extent of the Landers rupture on the northern segment (less than 10 km^{18,19,20}) constrain the depth of the transition from brittle to ductile behaviour to be ~10–15 km. Two distributed shear-strain models that are consistent with the 1 standard deviation (1 σ) uncertainties in the central Mojave data are depicted in Fig. 2. The slip rate in the models that fit the data (with $D=10$ –15 km, $W=30$ km) varies from 10 to 14 mm yr⁻¹. A slip rate estimate of about 12 mm yr⁻¹ across a 60-km-wide zone is consistent with geological data that suggest a slip rate over the past 5.5 Myr of 12 mm yr⁻¹ (ref. 6). Earlier geodetic estimates of the slip rate were lower, 7–8 mm yr⁻¹, and were obtained by simply multiplying the average strain rate by the width of the zone of strain accumulation, taken to be between 40 and 60 km (refs 7, 8).

In principle, the amount of time required to accumulate strain energy equivalent to that released in a given earthquake can be calculated by comparing the coseismic strain release with the rate of interseismic strain accumulation^{21,22}. We used angle changes measured between 1934 and 1982 at four sites that spanned the Landers rupture (Fig. 1) to estimate a rate of interseismic strain of $\dot{\gamma} = 0.22 \pm 0.09 \mu\text{rad yr}^{-1}$ (Table 1), a result not significantly higher than in adjacent segments of the shear zone. The average

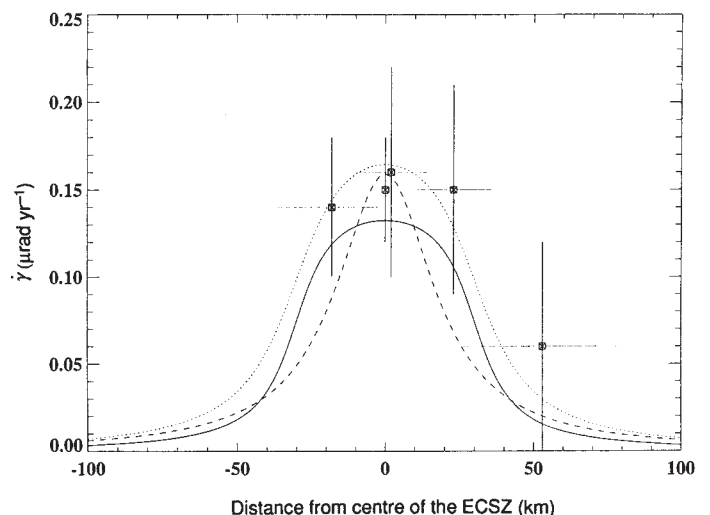


FIG. 2 Predicted and calculated shear-strain rate ($\dot{\gamma}$) plotted against distance from the centre of the eastern California shear zone (taken here to be half way between the Helendale and Pisgah faults); distances to the east and west are shown as positive and negative respectively. The shear-strain rate for the subnetworks and the average strain rate between the Helendale and Pisgah faults (Table 1) are given by squares; vertical error bars represent 1 σ . The approximate width of the four subregions is indicated by horizontal error bars. The predicted shear-strain rate for the distributed shear-strain model with a total width of 60 km ($W=30$) is given by the solid line for $D=10$ km and $\dot{b}=10$ mm yr⁻¹ and by the dotted line for $D=15$ km and $\dot{b}=14$ mm yr⁻¹. For comparison, the shear-strain rate predicted for 10 mm yr⁻¹ of slip on a single fault at a depth of 20 km (ref. 8) is given by the dashed line. The average strain rate across the Barstow trilateration network⁸, northeast of the central Mojave network, and spanning ~50 km is $0.13 \pm 0.02 \mu\text{strain yr}^{-1}$ at N33° W $\pm$ 4°.

shear-strain change across the Landers rupture between 1982 (when trilateration measurements were made) and July 1992 Global Positioning System measurements was $\sim 767 \mu\text{rad}$, more than a factor of 10 larger than the average coseismic strain drop for earthquakes of $\sim 50 \mu\text{rad}$ suggested by Rikitake²². Taking the ratio of the Landers coseismic strain drop ($767 \mu\text{rad}$) to the interseismic strain rate across the same region ($0.22 \mu\text{rad yr}^{-1}$) gives a strain recharge time of $\sim 3,500$ yr, or $\sim 5,000$ yr if we use the Camp Rock subnetwork (Table 1). The difficulty in using the strain energy recharge method to estimate a recurrence interval is that the size of the expected earthquake and the amount of interseismic strain associated with a single Mojave fault are uncertain. Independent palaeoseismic^{5,11,12} and seismic hazard studies¹⁵ suggest a recurrence time for individual earthquakes on Mojave faults of thousands to tens of thousands of years. In contrast, the time interval between the last 10 earthquakes on the Mojave segment of the San Andreas fault is between ~ 40 and 330 yr (ref. 23). Although the repeat time for individual earthquakes in the Mojave desert may be 10^3 or 10^4 yr, there are multiple faults that accommodate a slip rate that we estimate to be $\sim 1/3$ that of the adjacent San Andreas fault. Therefore, an earthquake would be expected somewhere within the central Mojave more frequently than the strain energy recharge rate for an individual event.

A long-term slip rate across the Mojave shear zone of $\sim 12 \text{ mm yr}^{-1}$ implies that the remaining relative plate motion of $\sim 34 \text{ mm yr}^{-1}$ (ref. 4) is accommodated by slip on, or to the west of, the San Andreas fault. Ground-based geodetic and VLBI data are consistent with this inference. From a dislocation model of the San Andreas fault in which the fault is locked below a depth of 20 km, ground-based geodetic data from the Mojave segment

imply a slip rate of $25 \pm 5 \text{ mm yr}^{-1}$ (ref. 24). A slip rate of $\sim 25 \text{ mm yr}^{-1}$ on this segment of the San Andreas fault, at the lower end of the generally cited range¹⁴, is also consistent with similar modelling of the difference velocity observed between VLBI sites PBLO and JPL (Fig. 1). $\square$

Received 27 April; accepted 12 November 1993.

- Jennings, C. W. Open-file Report No. 92-03 (Calif. Div. of Mines and Geol., Sacramento, 1992).
- Clark, M. M. et al. Open-file Report No. OFR-84-106 (US Geol. Surv., Reston, 1984).
- Lisowski, M., Savage, J. C. & Prescott, W. H. *J. geophys. Res.* **96**, 8369–8389 (1991).
- DeMets, C., Gordon, R. G., Argus, D. F. & Stein, S. *Geophys. J. Int.* **101**, 425–478 (1990).
- Wallace, R. E. *J. geophys. Res.* **89**, 5763–5769 (1984).
- Dokka, R. K. & Travis, C. J. *Geophys. Res. Lett.* **17**, 1323–1326 (1990).
- Sauber, J., Thatcher, W. & Solomon, S. C. *J. geophys. Res.* **91**, 12683–12693 (1986).
- Savage, J. S., Lisowski, M. & Prescott, W. H. *Geophys. Res. Lett.* **17**, 2113–2116 (1990).
- Sauber, J. thesis, Massachusetts Inst. of Technol. (1988); also as Technical memor. No. 100732 (NASA, Washington DC, 1989).
- Bartley, J. M., Glazner, A. F. & Schermer, E. R. *Science* **248**, 1398–1401 (1990).
- Lindvall, S. C. & Rockwell, T. K. *Geol. Soc. Am. Abstr. Prog.* **25**(5), 70 (1993).
- Cinti, F. R. et al. *Geol. Soc. Am. Abstr. Progr.* **25**(5), 21 (1993).
- Jones, L. M. *J. geophys. Res.* **93**, 8869–8891 (1988).
- Open File Report No. 88-398 (US Geol. Surv., Reston, 1988).
- Wesnosky, S. G. *J. geophys. Res.* **91**, 12587–12631 (1986).
- Prescott, W. H. & Nur, A. *J. geophys. Res.* **86**, 999–1004 (1981).
- Zhao, D. & Kanamori, H. *Geophys. Res. Lett.* **20**, 1083–1086 (1993).
- Murray, M. H., Savage, J. C., Lisowski, M. & Gross, W. K. *Geophys. Res. Lett.* **20**, 623–626 (1993).
- Blewitt, G. et al. *Nature* **361**, 340–342 (1993).
- Bock, Y. et al. *Nature* **361**, 337–339 (1993).
- Scholz, C. H. *The Mechanics of Earthquakes and Faulting* (Cambridge Univ. Press, 1990).
- Rikitake, T. *Earthquake Prediction* (Elsevier, Amsterdam, 1976).
- Sieh, K. E., Stuiver, M. & Brilling, D. J. *Geophys. Res.* **94**, 603–623 (1989).
- Eberhart-Phillips, D., Lisowski, M. & Zoback, M. D. *J. geophys. Res.* **95**, 1139–1153 (1990).
- Prescott, W. H. *Bull. Seism. Soc. Am.* **66**, 1847–1853 (1976).
- Seih, K. E. et al. *Science* **260**, 171–176 (1993).
- Ryan, J. W., Ma, C. & Caprette, D. S. Technical Memor. No. 104572 (NASA, Washington DC, 1993).

Sex ratio bias, relatedness asymmetry and queen mating frequency in ants

Liselotte Sundström

Department of Zoology, University of Helsinki, PO Box 17, FIN-00014 Helsinki, Finland

HAMILTON'S rule and the principle of inclusive fitness¹ provide a theoretical basis for understanding the evolution of social behaviour, and a framework for predicting reproductive characteristics of social insect colonies^{2–4}. Sex allocation in social insects (especially ants) has become a central factor in tests of inclusive fitness theory^{5–9}. The most powerful such test is the analysis of individual colonies where the predicted sex allocation varies depending on variation in worker fitness functions^{10,11}. Recently developed models^{5,12} predict that workers may enhance their inclusive fitness by biasing sex ratios in response to the degree of relatedness asymmetry in each colony. Here I provide the first empirical evidence of facultative sex ratio biasing in response to relatedness asymmetries caused by inter-colony variations in queen mating frequencies. In a Finnish population of the ant *Formica truncorum*, colonies have a single queen mated to one or several males. Colonies show a bimodal distribution of sex ratios, with a significantly greater proportion of males in colonies headed by a multiply mated queen.

The general prediction of inclusive fitness theory, that partial or complete worker control of sex allocation leads to overall female-biased sex ratios², agrees generally with empirical data^{6,7,13}. However, complications in the interpretation remain because a population level approach leaves the abundant variation in sex ratios among colonies unexplained¹⁴. Recent sex allo-

cation theory deals specifically with such inter-colony variation in sex ratios. Three mutually non-exclusive mechanisms have been proposed to explain colony level variation in sex ratios of social Hymenoptera. First, variations in resource levels may shift colony sex ratios towards male or female bias^{11,15}; second, differential dispersal of the sexes may result in different degrees of local mate competition or local resource competition^{16–18}, causing small colonies to specialize in one sex and large colonies to specialize in the other; and third, colonies with a high relatedness asymmetry are predicted to specialize in females and those with a lower relatedness asymmetry to specialize in males^{5,12,19}. (In Hymenoptera the haplodiploid sex-determining mechanism causes females to be on average more closely related to their sisters than to their brothers, compared with males; this is expressed as the relatedness asymmetry within a colony^{1,2,12}.) The first two hypotheses address factors associated with colony size and productivity, whereas the last model predicts an association between colony-specific relatedness structure and the sex ratio.

Observed sex-allocation patterns in halictid bees with eusocial and parasocial colonies^{20,21} and in some polygynous ants^{6,8} follow the predictions of relatedness-induced split sex ratio theory^{5,10,12}. In these cases, workers may assess the status of their colony by the presence or absence of a queen or by the presence of different numbers of functional queens. It has remained unclear, however, whether split sex ratio could work in cases where the variation in relatedness asymmetry is due solely to variation in queen mating frequencies. Here the assessment of colony type by investing workers would have to rely on some phenotypic expression of genetic diversity among nestmates, for example, genetically determined odour cues^{5,12,19,22}. Workers may control sex allocation both by selectively rearing or killing female and male larvae, and by nutritional control causing some female larvae to enter the queen development pathway while others develop into workers^{2,23,24}.

The number of matings per queen varies within many ant species^{25,26}. Colonies with one singly mated queen have a higher relatedness asymmetry (3:1) than colonies with one doubly mated (2:1) or one triply mated (1.67:1) queen. It follows from split sex ratio theory^{5,12} that if workers control the sex ratio they should raise largely or exclusively females in colonies with a singly mated queen and largely or exclusively males in colonies with a multiply mated queen. This study tests the theoretical prediction of an association between queen mating frequencies and sex allocation in an approximately panmictic population of the wood ant *F. truncorum* in the archipelago on the southwest coast of Finland. The 5-km² study area comprises five islands where 23 colonies of *F. truncorum* were found and sampled from 1989 to 1992.

The mating status of the queens in the study population was assessed using allozyme variation at four loci for both workers and males. The average intra-colony relatedness (r) of workers was 0.60 ± 0.05 (mean $\pm$ s.e.)²⁷, giving a population level relative relatedness asymmetry of 2.4:1 (0.60/0.25)¹². The intra-colony relatedness was also estimated separately for colonies categorized as either singly mated or multiply mated ($r_s = 0.72 \pm 0.04$, $r_M = 0.43 \pm 0.10$; t -test for unequal variances: $t = 2.7$, $P < 0.02$, d.f. = 21), the corresponding relatedness asymmetries being 2.9:1 for the singly mated class and 1.7:1 for the multiply mated class¹². This corresponds approximately to the values 3:1 and 2:1 used in the split sex ratio model by Boomsma and Grafen¹². The colonies persist for many years apparently without replacing the queen, and sex ratios in a given colony are stable over years (see Fig. 1 legend). The colony-level sex ratio data (expressed as proportional investment in females) agree well with the theoretical predictions: colonies with a multiply mated queen produce exclusively males or mixed sex ratios, whereas colonies with a singly mated queen produce only females or a mixed sex ratio (Fig. 1). Colony productivity also seems to increase female investment. When the effect of mating frequency

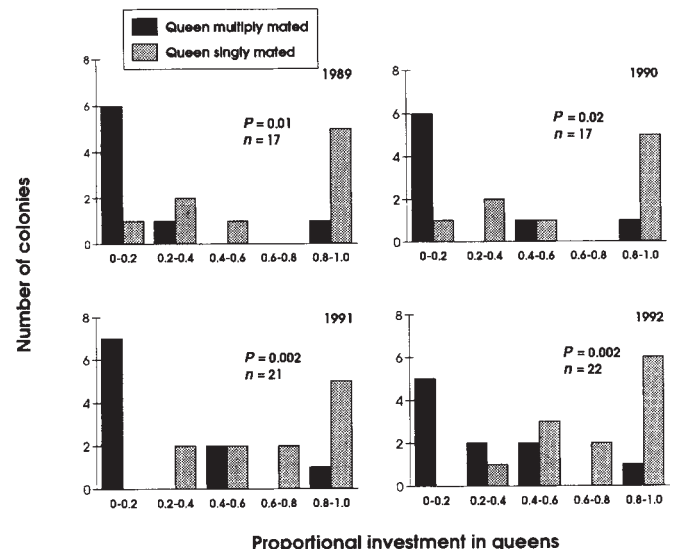
and productivity are considered simultaneously, there is virtually no overlap in sex ratios between the two colony categories (Fig. 2).

In cases where two males mated to a single queen have identical genotypes with respect to the diagnostic alleles, the colony queen would be misclassified as singly mated (see Fig. 1 legend). A reanalysis of the data with two colonies producing an all-female brood reassigned to the multiply mated category still produced significant differences in sex allocation between the two colony categories (see Fig. 2 legend). The results are thus robust even with this conservative approach.

The observed mating swarm sex ratio is 60% females (numerical ratio), and 67% when corrected for investment differences. This is slightly lower than the expected 71% females based on the observed 2.4:1 relatedness asymmetry in the whole population³. This measured relatedness asymmetry is independent of misclassifications and includes relative contribution of each colony type to the total population, as it is calculated directly from genetic data²⁷. The multiply mated class of colonies is expected to balance the population sex ratio to reflect its own relatedness asymmetry of 1.7:1 when its representation is more than 33% and the mean productivity of classes is equal¹². Indeed, the observed mating swarm sex ratio, corrected for sex-specific investment differences, equals the theoretical optimum for this class (63% females) instead of approximating the mean relatedness asymmetry of the whole population (71%). Partial queen control^{4,6,15,28} and differential dispersal of sexes causing local resource competition²⁹ cannot be fully excluded, but would simply cause an overall shift in total sex ratios. The prediction that colony sex ratios should be split in response to the colony-level relatedness asymmetry still remains. This also suggests that the population aspects of worker-controlled split sex ratios in *F. truncorum* agree with the predictions for full worker control¹², even after controlling for colony productivity.

The results thus strongly suggest that workers do control sex

FIG. 1 Sex ratios expressed as relative investment (dry weight) in queens in colonies having singly and multiply mated queens over a 4-yr period. All colonies examined produced brood each year. Five of the colonies were new in 1991, one in 1992 and one colony died in 1990, all other colonies being sampled for 4 yr. The social structure of the colonies was assessed each year by horizontal starch gel electrophoresis^{27,30} using allozyme variation at three polymorphic loci (*Idh*, *Pgk* and *Aco*). In 1993 another polymorphic locus (*Agpdh*) was added. For each colony, 10–40 workers and males were analysed and no colony showed any changes in genotypes over the years, indicating no queen turnover²⁷. The genotype distributions among male and worker offspring confirm that all males were queen produced²⁷. Allele frequencies were reasonably equal, making the assessment of matings reliable. Thirteen colonies were assessed as having a singly mated queen, eight as having a doubly mated queen and two as having a triply mated queen. The colonies with doubly and triply mated queens were combined in the further analyses. Based on the allele frequencies in the population, an estimated 13% of double matings would have passed undetected because of identical genotypes with respect to the diagnostic alleles among male mates of queens. Thus, one of the queens categorized as singly mated was likely to have been doubly mated. Final analyses were adjusted for this potential error (Fig. 2). Sexual pupae can be found in the nests for 2–3 weeks in summer and each colony was sampled during this period, with 1-week intervals, in each year (1989–92). The sex of pupae can be reliably determined immediately after pupation. Colony sex ratios were estimated from 20 randomly selected pupae in each sample and the arithmetic means of these samples were used. At eclosion the dry weights of females and males are equal (6.2 ± 0.3 mg and 6.1 ± 0.9 mg, respectively, mean $\pm$ s.d.; $t = 0.67$, $P > 0.05$). The dry weight of mature females was 8.75 ± 0.90 mg and that of males 6.60 ± 0.75 mg ($n = 40$ in both cases). The weights at maturity were used as estimates of the investment ratios. The sex ratio in a given colony is approximately constant over the years: of the total variance in sex ratios, 89% is due to differences among colonies, the variations among islands and among years being insignificant (two-level nested analysis of variance (ANOVA): between islands,



$F = 1.4$, $P > 0.05$; between colonies within island, $F = 85.2$, $P < 0.0001$; between years within colony, $F = 1.3$, $P > 0.05$). The sex ratios in colonies with multiply mated queens and with singly mated queens differed significantly, 53% of the total variance in sex ratios being due to differences in the number of matings among queens (two-level nested ANOVA: between colony types, $F = 16.4$, $P = 0.0008$; between colonies within colony type, $F = 0.8$, $P > 0.05$; between years within colonies, $F = 1.1$, $P > 0.05$), multiple mating being associated with male bias. These results also hold for each year separately, as indicated by the significance levels in the graphs (Mann-Whitney rank sum test).

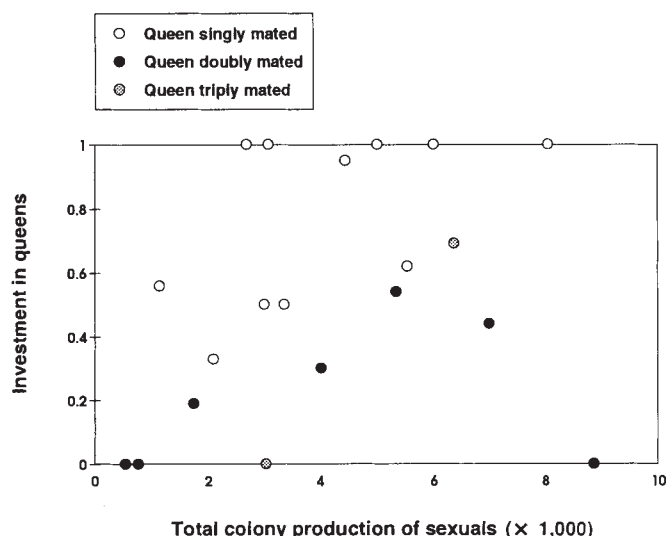


FIG. 2 Sex allocation in relation to colony productivity and queen mating frequency. The number of sexual pupae produced was determined twice by marking-recapturing in 1992. Marking was done with a waterproof filter pen and pupae were recaptured 2 days later. The amount of female bias depended mainly on colony type (singly or multiply mated queen), but also on colony productivity (ANOVA/analysis of covariance (ANCOVA): number of mates per queen, $F=26.7$, $P=0.0001$; with colony productivity as a covariate, $F=7.1$, $P=0.02$, $n=20$). Colony type and productivity together explain 66% ($F=17.1$, $P=0.0001$) of the total variation in sex ratios, and when productivity is factored out there is virtually no overlap in sex ratios between colony types. The association between colony productivity and sex ratio was confirmed by a non-parametric analysis on data adjusted for differences in means between the two categories (Spearman rank correlation $r=0.56$, $P=0.01$) as well as without this adjustment (Spearman rank correlation $r=0.43$, $P=0.04$). These results were checked for their robustness assuming that two of the colony queens assessed as singly mated were in fact multiply mated (see Fig. 1 legend). All possible pairs of colonies producing an all-female brood were selected and assigned to the multiply mated class. The ANOVA/ANCOVA above was repeated for each of the 10 pairs (ANOVA for number of mates per queen: the range of F values was 13.7–5.7, and the range of significance levels was $P=0.002$ –0.03; ANCOVA for productivity: $F=8.4$ –2.7, $P=0.01$ –0.12, $n=20$). In addition, the difference in sex ratios between the singly and multiply mated groups was tested with two randomly selected colonies producing an all-female brood reassigned to the multiply mated class and without productivity as a covariate (Mann-Whitney rank sum test, $P=0.03$). There were no differences in the total number of sexuals produced between colonies with singly mated and multiply mated queens ($4,030 \pm 2,000$ and $4,190 \pm 2,900$, respectively, Mann-Whitney, $P=1$).

ratios in response to the relative relatedness asymmetry in their colonies^{5,12}. This is a further example of the complexity and sophistication of worker reproductive strategies in social insects. It also implies that workers have a mechanism for accurately determining queen mating frequency, possibly using some phenotypic expression of the genetic diversity of their nestmates, such as genetically determined odour cues^{19,22}. The accuracy of the results further suggests that the diversity of such genetically determined cues is reasonably high^{19,22}. □

Received 20 August; accepted 4 November 1993.

1. Hamilton, W. D. *J. theor. Biol.* **7**, 1–52 (1964).
2. Trivers, R. L. & Hare, H. *Science* **191**, 249–263 (1976).
3. Benford, F. A. *J. theor. Biol.* **72**, 701–727 (1978).
4. Pamilo, P. *Am. Nat.* **137**, 83–107 (1991).
5. Boomsma, J. J. & Grafen, A. *Evolution* **44**, 1026–1034 (1990).
6. Nonacs, P. *Q. Rev. Biol.* **57**, 109–133 (1986).
7. van der Have, T. M., Boomsma, J. J. & Menken, S. B. *J. Evolution* **42**, 160–172 (1988).
8. Ward, P. S. *Behav. Ecol. Sociobiol.* **12**, 301–307 (1983).
9. Pamilo, P. *Behav. Ecol. Sociobiol.* **27**, 31–36 (1990).

10. Grafen, A. *J. theor. Biol.* **122**, 95–121 (1986).
11. Crozier, R. H. & Pamilo, P. In *Evolution and Diversity of Sex Ratio in Insects and Mites* (eds Wrensch, D. L. & Ebbert, M. A.) 369–383 (Chapman Hall, New York, 1992).
12. Boomsma, J. J. & Grafen, A. *J. evol. Biol.* **3**, 383–407 (1991).
13. Boomsma, J. J. *Am. Nat.* **133**, 517–532 (1989).
14. Alexander, R. D. & Sherman, P. W. *Science* **196**, 494–500 (1977).
15. Nonacs, P. *Evolution* **40**, 199–204 (1986).
16. Hamilton, W. D. *Science* **156**, 477–488 (1967).
17. Frank, S. A. *Theor. Populat. Biol.* **31**, 47–74 (1987).
18. Frank, S. A. *Behav. Ecol. Sociobiol.* **20**, 195–201 (1987).
19. Ratnieks, F. L. W. *Evolution* **45**, 281–292 (1991).
20. Boomsma, J. J. *Trends ecol. Evol.* **6**, 92–95 (1991).
21. Mueller, U. G. *Science* **254**, 442–444 (1991).
22. Ratnieks, F. L. W. *J. theor. Biol.* **142**, 87–93 (1990).
23. Brian, M. V. *Social Insects* (Chapman Hall, London, 1983).
24. Gosswald, K. *Die Waldameise* Vol. 1 (AULA, Wiesbaden, 1989).
25. Cole, B. J. *Behav. Ecol. Sociobiol.* **12**, 191–201 (1983).
26. Page, R. E. *Rev. Ent.* **31**, 297–320 (1986).
27. Sundström, L. *Behav. Ecol. Sociobiol.* (in the press).
28. Nonacs, P. In *Evolution and Diversity of Sex Ratio in Insects and Mites* (eds Wrensch, D. L. & Ebbert, M. A.) 384–401 (Chapman Hall, New York, 1992).
29. Clark, A. B. *Science* **201**, 163–165 (1978).
30. Seppä, P. *Behav. Ecol. Sociobiol.* **30**, 253–260 (1992).

ACKNOWLEDGEMENTS. I thank J. J. Boomsma, F. L. W. Ratnieks, P. Pamilo, R. Rosengren, K. Donner, A. Grafen, P. Seppä and W. Fortelius for comments on the manuscript, and S. Björkholm, L. Walin and C. Olsson for assistance in the field. The study was funded by the Academy of Finland.

Contrast dependence of colour and luminance motion mechanisms in human vision

Michael J. Hawken*, Karl R. Gegenfurtner*†‡ & Chao Tang*

* Center for Neural Science and † Howard Hughes Medical Institute, New York University, 4 Washington Place, New York, New York 10003, USA

CONVENTIONAL views of visual perception propose a colour-blind pathway conveying motion information and a motion-blind pathway carrying colour information^{1,2}. Recent studies show that motion perception is not always colour blind^{3,4}, is partially dependent on attention^{5,6}, can show considerable perceptual slowing around isoluminance^{7–9} and is contrast-dependent^{10,11}. If there is a single motion pathway, receiving luminance and chromatic input, then the dependence of relative perceived velocity on relative stimulus contrast should be the same for both luminance and chromatic targets. Here we provide a distinctive characterization of the motion mechanisms using a robust velocity-matching task. A relative contrast scale allows direct comparison of the performance with luminance and chromatic targets. The results show that the perceived speed of slowly moving coloured targets at isoluminance has a steep contrast dependence. The perceived speed of slowly moving luminance targets shows a much lower contrast dependence. At high speeds the contrast dependence is low for both luminance and isoluminant stimuli, although the behaviour is unlike either of the slow mechanisms. The results suggest two independent pathways that perceive slowly moving targets: one is luminance-sensitive and the other is colour-sensitive. Fast movement is signalled via a single motion pathway that is contrast-invariant and not colour blind.

We measured the perceived velocity of grating patterns of low spatial frequency as a function of contrast for slow (1 deg s⁻¹) and moderate velocities (8 deg s⁻¹). Gratings were either modulated in luminance or modulated along a red/green line in the isoluminant plane⁴. The slope of the function relating log-contrast to relative speed is invariant under multiplicative transformations of the contrast scale, and therefore allows a meaningful comparison between luminance and colour. Experienced

‡ Present address: Max-Planck-Institut für biologische Kybernetik, Tübingen, Germany.

psychophysical observers judged which of two spatially separated patterns moved faster. Figure 1a shows the difference in perceived velocity of five luminance gratings when compared with one of three standard luminance gratings. For the slowly moving standard, the lower-contrast gratings require a higher velocity to match the perceived speed of the higher-contrast gratings; this is consistent with previous reports^{10,11}. The slope of the function that relates perceived speed to contrast is the same across all standard contrasts, as indicated by the best-fitting regression lines in Fig. 1a.

When isoluminant red/green gratings are used as both the comparison and standard, the effect of contrast on perceived velocity is dramatically different from the results with luminance targets (Fig. 1b, triangles) under otherwise identical conditions. The dependence of perceived velocity on colour contrast is much stronger than for luminance contrast (Fig. 1b, squares compared with triangles). Similar results were obtained for isoluminant yellow/blue S-cone-isolating targets. The range of contrasts used for the isoluminant red/green gratings in our experiment went from the highest contrast we could attain with our monitor (corresponding to 10% and 24% cone contrast in the long- and middle-wavelength sensitive cones, respectively) down to 1 log unit below this. The luminance contrasts were selected to bracket the value of luminance contrast that produced a velocity match to the mid-point of the isoluminant contrasts. We have con-

firmed that there is little effect of spatial frequency; the same slope is seen for both square-wave and sinewave gratings.

Figure 1c shows the results for a series of experiments in which the speed of the standard grating target was 8 deg s^{-1} , otherwise the conditions were identical to those for the 1 deg s^{-1} experiments. There is a small but consistent increase in perceived speed with decreasing contrast for achromatic gratings (Fig. 1c, squares), confirming a previous report¹⁰. However, it is clear that, unlike the low-velocity data in Fig. 1b, at higher velocities the chromatically modulated targets behave exactly like the achromatic target.

To summarize the results across subjects, we have determined the relative slope for all the conditions. The results for four subjects are shown in Fig. 2, where there are three distinct clusters of slopes—one set of negative slopes (fast L and C conditions) associated with fast-moving targets, a second set of positive but shallow slopes (slow L) associated with the slow-moving luminance targets and a third group of steeper slopes (slow C) that are associated with the slow-moving chromatic targets.

For targets moving at moderate speeds, we show definitely that red/green isoluminant stimuli, which isolate the red/green colour opponent mechanism, are effective in producing a veridical percept of motion (Fig. 1c). Therefore, for these targets either the opponent colour system contributes effectively to

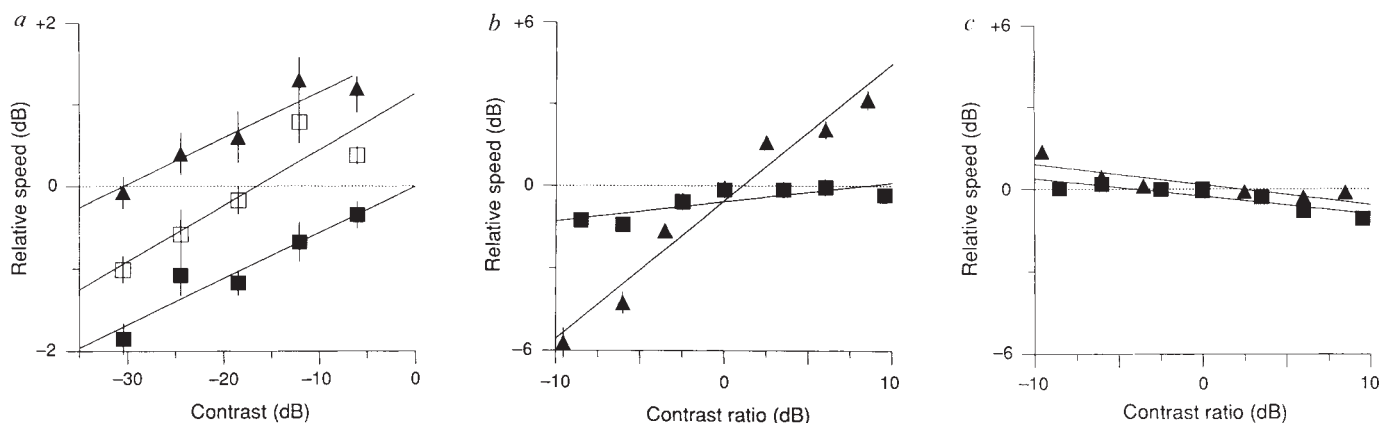


FIG. 1 The relationship between perceived velocity and contrast for different combinations of achromatic and chromatic sinewave gratings. All the data are for a single observer (mh); the slopes of each set of data for four observers are shown in Fig. 2. a, Each data set shows the difference in perceived velocity for five comparison gratings, each of different contrast, when compared with a single standard grating of 3, 12 or 50% contrast—triangles, open squares and closed squares, respectively. Contrast on the x-axis is given in decibels, 20 times the $\log_{10}$ of contrast. Relative speed (equal to velocity of standard pattern/velocity of comparison pattern at the match) is plotted on the y-axis. A positive value indicates that the actual velocity of the comparison was lower than the velocity of the standard when the subject judged the velocities as equal. A negative value means that the real velocity of the comparison was higher than the standard when the perceived velocities were equal. For this experiment both the standard and comparison were vertically oriented achromatic gratings with a spatial frequency of 1 cycle deg^{-1} . The speed of the standard was constant at 1 deg s^{-1} . To measure the thresholds, a spatial two-alternative forced-choice procedure was used to drive a staircase which adjusted the velocity of the comparison stimulus. The observer was instructed to fixate a small central square for the duration of the trial, which was 2 s. The next stimulus was presented immediately after the subject responded; no feedback was given. Stimuli were generated by a True-Vision Vista video controller and displayed on a Barco colour monitor (CCID 7351B), $28 \times 21 \text{ deg}$. The general methods of stimulus generation, calibration and control have recently been fully described elsewhere¹². Twelve pairs of turns on the staircase were found for each condition; each condition

was repeated six times. The same procedure was used throughout all the experiments. Subjects were seated 71 cm from a display monitor, positioned using a chin rest. They viewed binocularly with natural pupils. All subjects were optimally corrected to better than 20/20 and had normal colour vision. In a set of trials, there was a standard and a number of comparison stimuli, between 5 and 7. For a single trial, the standard was presented in either the upper or lower window (window size, $21^\circ \times 7^\circ$) and one of the comparisons was presented in the other window. The standard and the test drifted horizontally, in opposite directions. b, The standard and comparison gratings were either luminance/luminance (squares) or isoluminant/isoluminant (triangles). Otherwise, the conditions and experimental procedure are the same as described for a except for the contrast of the standards which are given below. In this graph and in c, contrast is expressed as the ratio of the contrast of the comparison to the contrast of the standard. Note that the relative velocity scale has been changed in this graph to accommodate the much greater slope of the isoluminant condition. c, Data for a faster velocity, 8 deg s^{-1} . The scale on the ordinate is the same as in b; note in particular that isoluminant gain (triangles) is much lower at the high velocity and indistinguishable from the luminance condition (squares). The contrast of luminance standard in b and c was 4%; the contrast of the isoluminant standard was 30% of the maximum attainable modulation. The contrasts were chosen to be approximately at equal multiples of detection threshold. Detection thresholds are -24 dB and -28 dB on the x-axis in b for luminance and colour, respectively. In c, the thresholds are -29 dB and -21 dB . The dotted lines in a, b and c show the prediction if velocity perception was contrast-independent.

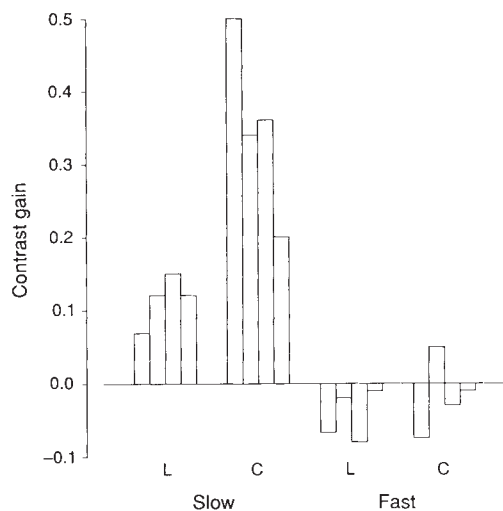


FIG. 2 The effect of contrast on the relative velocity of the comparison targets when compared with the standard target is given by the slope (contrast gain) of the regression lines in Fig. 1. This figure shows the contrast gain for four subjects in each of the four conditions. Slow and fast refer to the velocity of the standard targets, 1 deg s^{-1} and 8 deg s^{-1} , respectively. L and C refer to the two stimulus types, luminance and red/green isoluminant, respectively. For all four observers there is a substantially higher gain for the slow isoluminant stimulus. In a separate series of experiments, we measured both detection and identification threshold contours²⁰ of 1 cycle deg^{-1} drifting gratings for a range of temporal frequencies ranging from 1 to 16 cycles s^{-1} for a full range of directions in colour space which included points densely spaced around the isoluminant point. The isoluminant points were derived from these experiments.

motion perception or the luminance mechanism is sensitive to pure modulations in chromaticity. Whichever is correct, motion perception of moderate to fast movement is invariant to changes in contrast and chromaticity, such that the motion system is not blind to coloured targets.

The perception of slow movement is not invariant to changes in contrast and chromaticity. At low speeds there are two mechanisms for signalling the direction and velocity of moving targets. One is based on the well documented achromatic, directionally selective channel which is low pass in temporal frequency selectivity^{12,13}. The characteristics of the chromatically selective mechanism are not well known. One possibility is that the chromatic stimulus produces a response in the luminance mechanism. Origins of this signal could be chromatic aberration, variation in individual cells' isoluminant point or nonlinearities in the cells' response to contrast. These possibilities are ruled out by our findings. If the chromatic signal was operating by stimulating the neurons in the luminance pathway, it would be expected that the slope of the velocity-matching function for both red/green and yellow/blue targets would be identical to that of the luminance targets. This is clearly not the case.

Spatio-temporal energy filters are the basic building blocks for most current models of motion perception^{14,17}. The response of these filters confounds temporal frequency and contrast. To extract a correct velocity, most models of motion perception take the relative output of an array of these filters. Our results show that contrast invariance holds at moderate to high temporal frequencies, but that there are significant deviations from these models' predictions at low temporal frequencies.

We conclude that motion processing at moderate and fast speeds is independent of colour and proceeds via the established motion pathway which includes the extrastriate motion area MT. The neural mechanisms responsible for perceiving slowly moving targets are unclear. There is some evidence for the

involvement of the parvocellular pathway¹⁸. Our results indicate that at least two separate neural pathways are needed. □

Received 17 May; accepted 1 November 1993.

1. Zeki, S. M. *Nature* **274**, 423–428 (1978).
2. Livingstone, M. S. & Hubel, D. H. *J. Neurosci.* **7**, 3416–3468 (1987).
3. Cavanagh, P. & Anstis, S. *Vision Res.* **31**, 2109–2148 (1991).
4. Cavanagh, P. & Favreau, O. E. *Vision Res.* **25**, 1593–1601 (1985).
5. Cavanagh, P. *Science* **257**, 1563–1565 (1992).
6. Derrington, A. M. & Henning, B. G. *Vision Res.* **33**, 799–811 (1993).
7. Ramachandran, V. S. & Gregory, R. L. *Nature* **275**, 55–56 (1978).
8. Cavanagh, P., Tyler, C. W. & Favreau, O. E. *J. opt. Soc. Am. A*, **1**, 893–899 (1984).
9. Mullen, K. & Bolton, J. C. *Vision Res.* **32**, 483–488 (1992).
10. Thompson, P. *Vision Res.* **22**, 377–380 (1982).
11. Stone, L. S. & Thompson, P. *Vision Res.* **32**, 1535–1549 (1992).
12. Kulikowski, J. J. & Tolhurst, D. J. *J. Physiol., Lond.* **232**, 149–163 (1972).
13. Watson, A. B. & Robson, J. G. *Vision Res.* **21**, 1115–1122 (1981).
14. Adelson, E. H. & Bergen, J. R. *J. opt. Soc. Am. A*, **2**, 284–299 (1985).
15. Watson, A. B. & Ahumada, A. J. *J. opt. Soc. Am. A*, **2**, 322–342 (1985).
16. van Santen, J. P. H. & Sperling, G. *J. opt. Soc. Am. A*, **2**, 300–321 (1985).
17. Heeger, D. J. *J. opt. Soc. Am. A*, **4**, 1455–1471 (1987).
18. Merigan, W. H., Byrne, C. E. & Maunsell, J. H. R. *J. Neurosci.* **11**, 3422–3429 (1991).
19. Gegenfurtner, K. R. & Kiper, D. C. *J. opt. Soc. Am. A*, **9**, 1880–1888 (1992).
20. Stromeyer, C. F. III, Cole, G. R. & Kronauer, R. E. *Vision Res.* **25**, 219–237 (1985).

ACKNOWLEDGEMENTS. We thank C. Stromeyer for comments on an earlier version of this paper.

Oscillation and noise determine signal transduction in shark multimodal sensory cells

Hans Albert Braun, Heimo Wissing*, Klaus Schäfer* & Martin Christian Hirsch

Institute of Physiology, University of Marburg, Deutschhausstrasse 2, D-35037 Marburg, Germany

OSCILLATING membrane potentials that generate rhythmic impulse patterns are considered to be of particular significance for neuronal information processing^{1–4}. In contrast, noise is usually seen as a disturbance which limits the accuracy of information transfer^{5–8}. We show here, however, that noise in combination with intrinsic oscillations can provide neurons with particular encoding properties, a discovery we made when recording from single electro-sensory afferents of a fish. The temporal sequence of the impulse trains indicates oscillations that operate near the spike-triggering threshold. The oscillation frequency determines the basic rhythm of impulse generation, but whether or not an impulse is actually triggered essentially depends on superimposed noise. The probability of impulse generation can be altered considerably by minor modifications of oscillation baseline and amplitude, which may underlie the exquisite sensitivity of these receptors to thermal and electrical stimuli. Additionally, thermal, but not electrical, stimuli alter the oscillation frequency, allowing dual sensory messages to be conveyed in a single spike train. These findings demonstrate novel properties of sensory transduction which may be relevant for neuronal signalling in general.

Oscillating spike generation in sensory afferents is primarily indicated by a rhythmic grouping of impulses called 'bursts'^{9–12}. However, there is a different type of impulse pattern (Fig. 1) which also implies oscillating transduction processes but which, additionally, has an essential noise component^{11–15}. The oscillation does not trigger groups of impulses but merely initiates single spikes, with occasionally no impulse at all (Fig. 1a). Despite the seemingly irregular impulse trains, the interval histograms show distinct modes (Fig. 1b): the interspike intervals

* Present addresses: Departments of Anaesthesiology (H.W.) and Physiology (K.S.), University of Frankfurt, Theodor-Stern-Kai 7, D-60596 Frankfurt, Germany.

are concentrated at about integer multiples of a basic interval duration. The first distribution peak corresponds to the oscillation period, whereas the additional distribution peaks, comprising longer intervals, indicate the occurrence of subthreshold oscillations.

Identical patterns were described almost fifty years ago¹⁶, but attracted little subsequent attention and were sometimes described as 'abnormal' discharges¹⁷. More recently, it has become established that intrinsic oscillations underlie thermoreceptor function^{11, 13, 18, 19} and also seem to be involved in mechanoreception²⁰. The effects of noise in neuronal transduction, for example, stochastic resonance phenomena, have been addressed by mathematical modelling^{21, 22} and have recently attracted particular attention as a key factor for the generation of rhythmic impulse patterns in auditory, visual and mechanosensory neurons^{14, 15, 23}. However, the rhythmicity of the discharge in these and many other examples²⁴ was not intrinsic, but followed a periodic stimulus input. Here we present experimental evidence that intrinsic oscillations *per se* have an information-carrying role and in combination with noise can signal both environmental changes and modality-specific information.

The data were obtained from dogfish ampullae of Lorenzini, which are secondary sensory afferents located on the jaws of dogfish and other elasmobranchs²⁵ (Fig. 2a, b), thought to be important in prey detection²⁶. The ampullae are generally considered to be electroreceptors, but respond to thermal stimuli as well^{27, 28}.

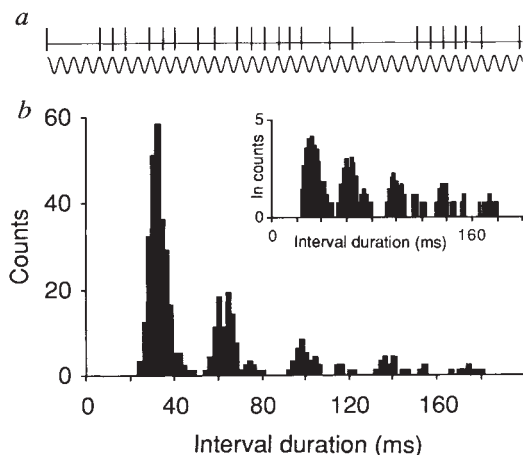


FIG. 1 Spike train with multimodal interval distribution recorded from a sensory afferent. **a**, Impulse sequence (upper trace) of 1,200 ms with normalized impulse height. The sine wave (lower trace) illustrates the rhythm of spike generation and the occasional occurrence of subthreshold oscillations. It was adjusted to the spike train by a modified cross-correlation algorithm. The sine wave was used as the basic periodic function; this does not mean that the hypothetical membrane potential oscillation is actually sinusoidal. **b**, Interspike-interval histogram from the recording shown in **a** (50 s, 500 intervals, bin width 3 ms; inset has a logarithmically scaled ordinate). The distribution of interspike intervals at integral multiples of a basic discharge period (~33 ms corresponds to 30 Hz) indicates that the impulses are generated by an oscillating process which is superimposed by noise. The first distribution peak corresponds to the oscillation period, and the following peaks indicate subthreshold oscillations, that is, each oscillation cycle without impulse triggering lengthens the interval duration by the oscillation period. The logarithmic decay of the histogram peaks (see inset) supports the assumption of a stochastic nature of the 'noisy' components¹⁴. The spikes were recorded from a spontaneously active fibre of a dogfish ampulla of Lorenzini²⁵ (see Fig. 2) at 18 °C. Methods used were as for Fig. 2.

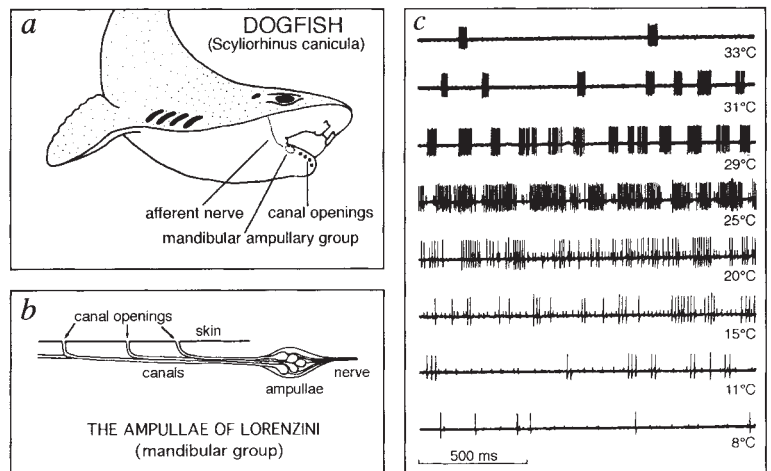


FIG. 2 Schematic drawings of location (**a**) and structure (**b**) of the mandibular group of the ampullae of Lorenzini²⁵ of the dogfish (*Scyliorhinus canicula*). **c**, Original recordings of afferent activity at various constant temperatures (same fibre as in Fig. 3). Spike amplitude was 220 μ V.

METHODS. The experiments were performed on isolated preparations. The fish was decapitated and the mandibular ampullary group removed. The ampullary canals were about 5 mm, the afferent nerve 30–40 mm long. The preparation was placed on a thermode and superfused with selachian saline solution³⁰. The thermode provided exactly controlled temperature stimuli through a custom-made thermostimulator. Impulse activity of single fibres ($n=120$) was recorded extracellularly after microdissection of the afferent nerve. A spike-level discriminator was used to separate single-fibre activity from background discharge and noise.

The discharge patterns recorded from a single ampullary nerve fibre at various constant temperatures differ considerably (Fig. 2c). The interspike-interval histograms (Fig. 3a) show a major distribution peak for each temperature which represents the oscillation period. As the temperature increases, the peak shifts continuously to shorter values, indicating an acceleration of the oscillation. There is a parallel increase in firing rate (Fig. 3b), but above a certain temperature the firing rate declines in parallel with an increased incidence of longer interspike intervals (small distribution components in Fig. 3a). This results from an increasing number of subthreshold oscillations, that is, the probability of impulse generation per oscillation cycle decreases.

These data (Fig. 3) demonstrate that a stationary temperature is unambiguously encoded in the oscillation frequency but not in the average firing rate. The information provided by the latter is ambiguous because it is determined not only by the oscillation frequency but also by the probability of impulse generation which is linked to the noise component. Therefore, with respect to the stationary mean firing rate, noise can be seen as a disturbance. Noise, however, is essential for the detection of electrical stimuli and the rate of temperature changes. The following data show that the signalling of these stimuli is based on modifications of the probability of impulse generation.

We have analysed the temporal sequences of successive interspike intervals during ramp-like stimuli (Fig. 4). In these plots (Fig. 4c), a band of short intervals represents the period of the spike-triggering oscillations. In response to a warming ramp, the intervals are gradually shortened in the manner expected from the stationary temperature dependencies of the interspike-interval histograms (Fig. 3a). In contrast, the number of subthreshold oscillations exhibits a transient response: warming induces a broad scattering of longer intervals which subsequently converge to form distinct bands (Fig. 4c). The firing rate (Fig. 4b) is transiently reduced in spite of the acceleration of the oscillation. Cooling ramps (not shown) produce the opposite effects: a gradual slowing of the oscillation frequency which is superimposed by a transient reduction of subthreshold oscillations.

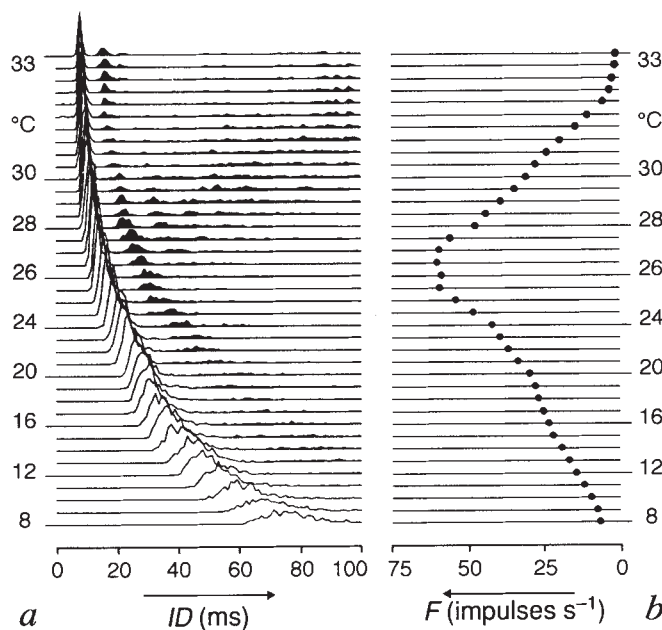


FIG. 3 Interval histograms (a) and corresponding firing rate (b) of a single ampullary afferent illustrate stationary temperature dependencies. a, Each histogram represents 500 intervals, bin width 1 ms; base-line distance between histograms corresponds to 25 intervals. a, The interval histograms (ID , interval duration) exhibit a major distribution peak which represents the oscillation period and shifts continuously from 71 ms at 8 °C to 7 ms at 33 °C. Additional distribution peaks (black shading) represent longer intervals which indicate subthreshold oscillations. At temperatures above 26 °C these distribution peaks are considerably scattered to higher values. Intervals longer than 100 ms are not shown. b, The firing rate (F) increases from 6 impulses s^{-1} at 8 °C to a maximum value of 62 impulses s^{-1} at ~26 °C and declines to 2 impulses s^{-1} at 33 °C. The corresponding oscillation frequencies are 14, 81 and 143 Hz.

METHODS. The figures were drawn from a continuous record of quasi-stationary impulse discharge ($n=63,000$ impulses) covering the whole activity range. Quasi-stationary conditions were obtained by a temperature increase of 0.006 °C s^{-1} , which was too slow to induce any dynamic effect (no hysteresis in the frequency-temperature relation was observed when temperature ramps of opposite direction were applied). The firing rate (b) was calculated directly from the corresponding interspike-interval histograms (a) and was plotted parallel to them. The temperature scale above 24 °C was expanded.

These results show that whereas stationary temperatures are principally encoded in the oscillation period, the dynamic responses are also governed by the probability of impulse generation.

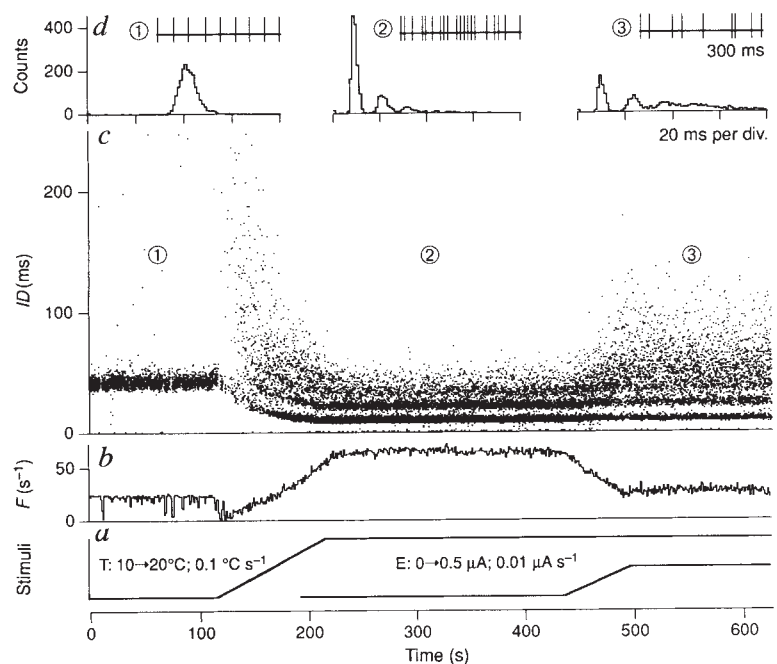
Electrical stimuli have no effect on the oscillation period but exclusively change the probability of impulse generation. Anodal currents reduce the number of shorter intervals in favour of a broad scattering of longer intervals (Fig. 4c, d), whereas cathodal currents concentrate the intervals at shorter values (not shown). The position of the distribution bands of interspike intervals remains the same (Fig. 4c), that is, the oscillation frequency is unaffected. The firing rate (Fig. 4b) simply parallels the varying occurrence of subthreshold oscillations.

As the oscillation frequency is unaffected by electrical stimuli, the ampullae of Lorenzini clearly distinguish between electromagnetic fields and environmental temperature changes. These differential encoding properties are manifested in the impulse pattern, but not necessarily in the averaged impulse rate. Thus, under different stimulus conditions the firing rate may be the same, whereas the activity pattern is considerably changed (for example, see beginning and end of the recording in Fig. 4).

Our findings clearly demonstrate that single sensory afferents of the ampullae of Lorenzini can signal modality-specific information in response to different stimuli. The oscillation frequency encodes absolute temperatures, whereas changes in the probability of impulse generation provide additional information about

FIG. 4 Discharge characteristics of ampullary receptors in response to ramp-shaped thermal and electrical stimulation. a, Stimuli (T , temperature; E , electrical current). b, Firing rate (F) plotted as peri-stimulus-time histogram (time base 1 s). c, Plot of successive interspike-intervals (ID , interval duration, $n=20,000$ intervals). d, Spike sequence and interval histogram patterns under different stimulus conditions (indicated by encircled numbers in the ID plot). At 10 °C (no electrical stimulus) the fibre discharged a regular train of spikes which was interrupted only occasionally by longer intervals. These discharge interruptions are not seen in the spike sequence and histogram (d,1) but are indicated in the ID plot (c) and in the firing rate (b). The temperature ramp from 10 to 20 °C gradually shortened the initial discharge period from about 40 ms to about 10 ms. Simultaneously, the number of longer intervals increased before they converged to form distinct bands (c,2; d,2). The discharge frequency was reduced at the beginning of the ramp but subsequently adapted to values higher than those initially obtained. The electrical ramp had no effect on the position of the bands of intervals but reduced the number of shorter intervals in favour of a broad distribution of longer intervals. Nevertheless, at least three different bands (c,3) or peaks (d,3) can still be recognized. The firing rate was reduced to almost the same level as that at the beginning of this recording, but the impulse pattern has changed considerably in accordance with the stimulus conditions.

METHODS. Electrical currents were applied via glass micropipettes which were inserted into the ampullary canal. The micropipettes were filled with saline solution and connected to the current source by a



platinum wire. For the current-controlled stimulation, the voltage was measured at a high series resistance. The stimulus function was provided by a ramp generator. For thermal stimulation, see Fig. 2 legend.

the rate of temperature change. Alterations of afferent activity without concomitant changes in the oscillation frequency unequivocally signal electrical stimuli.

These differential encoding properties do not need stimulus-dependent modifications of noise, even though it is an essential component. On the one hand, without noise, these electroreceptors would completely fail to detect electrical stimuli. On the other hand, when noise is added to internal oscillations, the probability of spike generation might be considerably altered by only small depolarizing or hyperpolarizing shifts of the oscillation baseline. This might explain the remarkable sensitivity of sharks to weak electromagnetic fields.

As the electrosensitivity of ampullary receptors is attributed to the presynaptic cells²⁸, we assume that the probability of spike generation is a function of stimulus-dependent transmitter

release. Modifications of the oscillation period, which also occur on temperature stimuli, might be attributed to direct effects on the oscillating membrane mechanisms presumably located in the postsynaptic cells.

It can be speculated that the transduction mechanisms described here may occur wherever intrinsic oscillations that operate near the spike-triggering threshold are subjected to different kinds of external modulations. These may be caused by physical stimuli, but may also be caused by neurotransmitters or neuromodulators. In fact, related patterns have recently been observed in mechanosensory cells²⁰ and certain cortical neurons²⁹. Therefore, similar mechanisms to those described here could well operate not only in sensory transduction but also in information processing in the central nervous system. □

Received 12 May; accepted 25 October 1993.

1. Linas, R. R. *Science* **242**, 1654–1664 (1988).
2. Aldenhoff, J. B. *Pharmacopsychiatry* **22**, 227–240 (1989).
3. Connors, W. B. & Gutnick, M. J. *Trends neuro. Sci.* **13**, 99–104 (1990).
4. McGormick, D. A. & Feese, H. R. *Neuroscience* **39**, 103–113 (1990).
5. Bullock, T. H. *J. gen. Physiol.* **55**, 565–584 (1970).
6. Fitzhugh, R. J. *gen. Physiol.* **40**, 675–692 (1958).
7. Schreiner, R. C., Essick, G. K. & Whitsel, B. L. *J. Neurophysiol.* **41**, 338–349 (1978).
8. Bialek, W. & Rieke, F. *Trends neuro. Sci.* **15**, 428–435 (1992).
9. Iggo, A. *J. Physiol., Lond.* **200**, 391–407 (1969).
10. Dykes, R. W. *Brain Res.* **98**, 485–500 (1975).
11. Braun, H. A., Bade, H. & Hensel, H. *Pflügers Arch.* **386**, 1–9 (1980).
12. Braun, H. A., Schäfer, K. & Wissing, H. in *Thermoreception and Temperature Regulation* (eds Bligh, J. & Voigt, K.) 19–29 (Springer, Berlin, 1990).
13. Schäfer, K., Braun, H. A. & Rempe, L. *Experientia* **47**, 47–50 (1990).
14. Longtin, A., Bulsara, A. & Moss, F. *Phys. Rev. Lett.* **67**, 656–659 (1991).
15. Maddox, J. *Nature* **325**, 469 (1991).
16. Brink, F., Bronk, D. W. & Larabee, M. G. *Ann. N.Y. Acad. Sci.* **47**, 457–485 (1946).
17. Buller, A. J., Nicholls, J. G. & Ström, G. *J. Physiol., Lond.* **122**, 409–418 (1953).

18. Braun, H. A., Schäfer, K., Wissing, H. & Hensel, H. in *Sensory Receptor Mechanisms* (eds Hamann, W. & Iggo, A.) 147–156 (World Science, Singapore, 1984).
19. Schäfer, K. & Braun, H. A. *Pflügers Arch.* **417**, 91–99 (1991).
20. Sokabe, M. et al. *J. Neurophysiol.* **70**, 275–283 (1993).
21. Knight, B. W. *J. gen. Physiol.* **59**, 734–766 (1972).
22. Longtin, A. *J. stat. Phys.* **70**, 309–327 (1993).
23. Douglass, J. K., Wilkens, L., Pantazidou, E. & Moss, F. *Nature* **365**, 337–340 (1993).
24. Fohmeister, J., Poppele, R. E. & Purple, R. L. *J. Neurophysiol.* **37**, 1213–1227 (1974).
25. Waltman, B. *Acta physiol. scand.* **66**, Suppl. 264, 1–60 (1966).
26. Kalmijn, A. *Nature* **212**, 1232–1233 (1966).
27. Bromm, B., Hensel, H. J. & Nier, K. *Experientia* **31**, 615–618 (1975).
28. Akoev, G. N. & Andrianov, G. N. *Adv. physiol. Sci.* **31**, 57–73 (1980).
29. Alonso, A. & Klink, R. *J. Neurophysiol.* **70**, 128–143 (1993).
30. Hensel, H. *Z. vergl. Physiol.* **37**, 509–526 (1955).

ACKNOWLEDGMENTS: We thank G. McGregor and C. Walther for advice and discussions, K. Voigt for encouragement, and the Biologische Anstalt Helgoland for providing their facilities. This work was supported by the Deutsche Forschungsgemeinschaft.

Calcium controls light-triggered formation of catalytically active rhodopsin

Leon Lagnado* & Denis A. Baylor

Department of Neurobiology, Sherman Fairchild Science Building, Stanford University School of Medicine, Stanford, California 94305, USA

BACKGROUND light reduces the gain of phototransduction in retinal rods so that the ability to register changes in light intensity is not prevented by saturation of the cell's response^{1,2}. The gain is reduced by a light-induced fall in the intracellular calcium concentration which results from blockage of Ca^{2+} entry through the channels closed by light and continued Ca^{2+} extrusion by the Na:Ca,K exchanger^{3–7}. Calcium seems to exert several coordinated effects on the cyclic GMP cascade: a fall in $[\text{Ca}^{2+}]$ stimulates cGMP synthesis^{8,9}, increases the affinity of the cGMP-gated channel for cGMP¹⁰ and accelerates rhodopsin deactivation by phosphorylation¹¹. We now report that lowering intracellular $[\text{Ca}^{2+}]$ reduces the catalytic rhodopsin activity produced by light. The effect is operationally equivalent to a fourfold reduction in the number of rhodopsin molecules available for activation. The reduction in gain is cooperative and half-maximal at about 35 nM Ca^{2+} , suggesting that it is mediated by a specific Ca^{2+} -binding protein. Reduced rhodopsin activity in low Ca^{2+} should contribute to adaptation in background light.

We examined the effects of Ca^{2+} on the phototransduction cascade by recording the membrane current from rod outer segments of the tiger salamander, *Ambystoma tigrinum*, that had been opened by mechanical truncation and internally dialysed

(Fig. 1a). In this preparation the intracellular concentrations of Ca^{2+} and other small molecules can be rapidly changed. Figure 1b shows that lowering the free calcium ion concentration ($[\text{Ca}^{2+}]$) inside the outer segment from 1 μM to 30 nM reduced the response to dim steady light. When the $[\text{Ca}^{2+}]$ was restored to 1 μM the light response increased. The antagonism of the light response in low Ca^{2+} resulted from reopening of cGMP-activated channels. This was not due to stimulation of guanylate cyclase because lowering $[\text{Ca}^{2+}]$ in darkness immediately after the light exposure had little effect on the membrane current, even though channel activation by cGMP was far from saturation. This indicates that cyclase activity was negligible with only 10 μM GTP in the dialysis solution (Fig. 1 legend). The decreased light response in low Ca^{2+} can therefore be attributed wholly to inhibition of light-stimulated cyclic GMP-phosphodiesterase (PDE) activity. We analysed this effect and found that it results from reduced formation of active rhodopsin.

Figure 1c shows families of responses to brief flashes of increasing strength in 1 μM Ca^{2+} and 0 Ca^{2+} . The relations between response amplitude and flash strength in Fig. 1d were fitted by saturating exponential functions¹². Lowering $[\text{Ca}^{2+}]$ shifted the relation to the right, an effect equivalent to dimming the flashes by a fixed factor. In an intact rod, background light also shifts the relation to the right¹³. At the start of the experiment in Fig. 1d, the half-saturating flash strength ($I_{1/2}$) was 4.4 photons μm^{-2} , similar to that of an intact salamander rod. In 0 Ca^{2+} $I_{1/2}$ was 21.5 photons μm^{-2} , so that the sensitivity was fivefold lower. Sensitivity increased, but was not fully regained, on returning to 1 μM Ca^{2+} ($I_{1/2} = 7.1$ photons μm^{-2}). The Ca^{2+} dependence of the response amplitude often disappeared when a cell was dialysed for more than 15 min, suggesting that it depended on soluble factors that were gradually lost.

Removing Ca^{2+} reduced the rising phase of the flash response without affecting the recovery phase. The rising phases of the flash family in 1 μM Ca^{2+} (Fig. 1c) are plotted on an expanded timescale in the left panel of Fig. 2a. The responses rose more

* Present address: MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK.

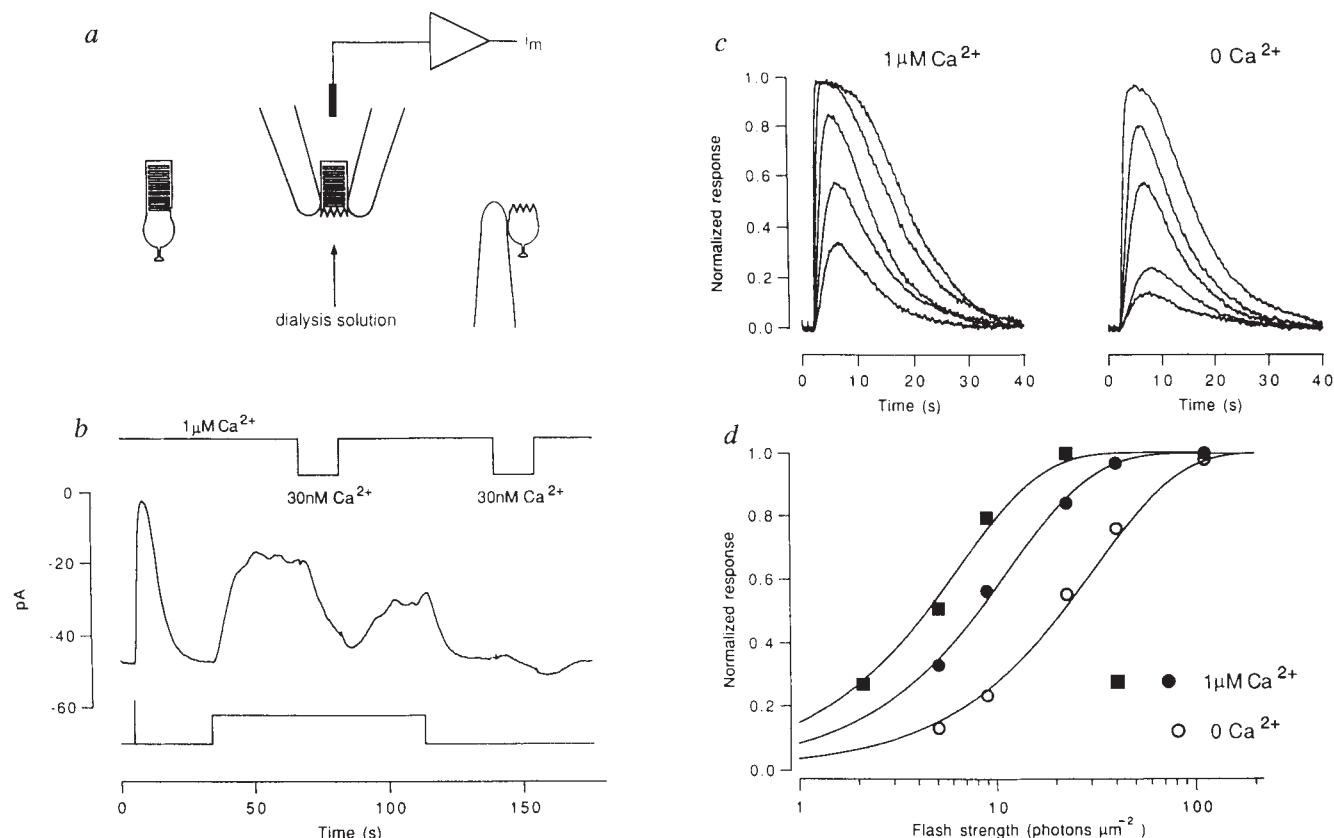


FIG. 1 Lowering $[Ca^{2+}]$ reduced light-dependent PDE activity in internally dialysed salamander rod outer segments. **a**, Truncation of a rod. An intact rod (left) and a truncated rod outer segment held in a suction electrode (right) which collects its membrane current. The glass probe has amputated the inner segment. **b**, Membrane current recording showing a reduction in the response to dim, steady light when the free $[Ca^{2+}]$ in the dialysing solution was lowered from $1 \mu M$ to $30 nM$. Timing of light shown by trace below current recording; timing of solution change shown above. Flash delivered $43 \text{ photons } \mu m^{-2}$ and step delivered $2.8 \text{ photons } \mu m^{-2} s^{-1}$. Similar results were observed in 35 other cells. Both exposures to $30 nM Ca^{2+}$ lasted 15 s. Note that the second exposure, in darkness, activated only a small inward current, indicating that cGMP synthesis was negligible when $[Ca^{2+}]$ was lowered in $10 \mu M$ GTP. The small increase in the inward current also occurred in the absence of GTP and represents an increase in the affinity of the cGMP-gated channels¹⁰. **c**, Superimposed responses to 20-ms flashes delivered at 2 s, in $1 \mu M Ca^{2+}$ (left) and $0 Ca^{2+}$ (right). Dimmest flash delivered $5 \text{ photons } \mu m^{-2}$, the brightest $113 \text{ photons } \mu m^{-2}$. All responses were normalized by the dark current. Flashes in $0 Ca^{2+}$ were started 6 min after truncation and those in $1 \mu M Ca^{2+}$ after 11 min. Dialysing solution contained $50 \mu M$ ATP (see Methods). **d**, Peak amplitude of the normalized response plotted against flash strength (log scale). The curves were drawn according to the function $r = 1 - \exp(-kl)$, where r is the normalized response, l is the flash strength and k is a constant related to the half-saturating flash strength $l_{1/2}$ by $k = (\ln 2/l_{1/2})$. Circles are from traces in **c**. Filled squares are from a run in $1 \mu M Ca^{2+}$ 1.3 min after truncation. Values of $l_{1/2}$ were (in $\text{photons } \mu m^{-2}$): 4.4 ($1 \mu M Ca^{2+}$), 21.5 ($0 Ca^{2+}$) and 7.1 (return to $1 \mu M Ca^{2+}$); corresponding values of dark current were 27, 52 and 39 pA. The sensitivity fell by a factor of 5 on removing Ca^{2+} , and increased by a factor of 3 on returning to $1 \mu M Ca^{2+}$.

METHODS. Suction electrodes were used to record membrane current from truncated outer segments²² of rods from the larval tiger salamander (**a**). Methods for obtaining isolated rods by mechanical dissociation of the retina, delivering light stimuli, electrical recording and rapid switching of solutions have been described^{23,24}. Light flashes were at a wavelength of 500 nm (bandwidth at half-height 10 nm). Truncations were done in a solution containing $0.1\text{--}1 \text{ mM}$ cGMP. When the interior of the outer segment was exposed to the bathing solution a large inward current rapidly developed as cGMP diffused in through the cut end. Cells were isolated in Ringer solution containing (in mM): NaCl, 110; $CaCl_2$, 1; $MgCl_2$, 1.6; KCl, 2.5; HEPES, 10; pH 7.6 with TMAOH (tetramethyl ammonium hydroxide). The suction electrode was filled with a

similar solution without KCl and $CaCl_2$ and with $200 \mu M$ EGTA added to buffer the free calcium to below 5 nM. The cell interior was dialysed with a solution containing (in mM): arginine glutamate, 110; $MgCl_2$, 2; HEPES, 5; EGTA, 0.2; $CaCl_2$, 0.197; cGMP, 0.1; ATP, 1; GTP, 0.01; pH 7.7 with TMAOH; calculated free $[Ca^{2+}] = 1 \mu M$. In later experiments 1 mM EGTA was used. The final free $[Ca^{2+}]$ was set by adding an appropriate amount of a 1 M $CaCl_2$ stock, calculated as described elsewhere²⁵; GTP (Li^+ salt) ATP (Na^+ salt) and cGMP (Na^+ salt) were from Sigma. Arginine is impermeant in the cGMP-activated channel, and the inward current activated by cGMP was carried by sodium in the suction electrode entering the outer segment. An inward sodium gradient was established by the low ($\leq 2 \text{ mM}$) sodium concentration in the dialysing solution. Membrane potential was held at 0 mV. $Na:Ca,K$ exchange was blocked by omitting Ca^{2+} from the external solution and omitting K^+ from both intra- and extracellular solutions. Under these conditions, the only Ca^{2+} flux will be a small outward leak through the light-sensitive channel. So as to use the cGMP-activated current to assay changes in PDE activity, it was necessary to minimize guanylate cyclase activity by lowering the [GTP] in the dialysing solution to $10 \mu M$ (see text). To allow the time course of PDE activity after a flash to be assessed, the cGMP-activated current in darkness was kept well below saturation by dialysing with $100 \mu M$ cGMP; this caused 10–30% of the cGMP-activated channels to be open²². Including ATP in the dialysing solution allowed rhodopsin to be shut off by rhodopsin kinase²⁶. Recovery of the dark current requires that cGMP diffuse back into the outer segment from the external solution, but this was not rate-limiting for recovery. The rate of recovery after a flash was significantly slower than the rate of activation of the dark current on applying $100 \mu M$ cGMP (1.55 ± 0.2 -fold slower; mean $\pm$ s.d.; 13 measurements in 4 cells). The rate of cGMP diffusion was not affected by the presence or absence of Ca^{2+} or nucleotides in the dialysis solution. The rate of recovery after a flash is limited by R^* deactivation in the presence of recoverin (L.L., S. Zozulya, L. Stryer and D.A.B., unpublished observations), but recoverin was absent in these experiments. In the experiment shown in Figs 1, 2 and 4a, we ensured that rhodopsin phosphorylation was rate-limiting for recovery by using $50 \mu M$ ATP. In this solution the response fell with a time-constant of 9 s, whereas in 1 mM ATP it was 5 s. Low $[Ca^{2+}]$ also desensitized responses obtained in 1 mM ATP without affecting the rate of recovery. Even in 1 mM ATP, flash responses generally lasted longer than those of an intact rod, probably because the free $[Ca^{2+}]$ did not fall during the response²⁷. When the recovery phase of the flash response slowed appreciably, indicating a long-term change in the state of the cell, the experiment was stopped.

rapidly with increasing flash strength because PDE activation was more intense¹⁴. At early times after a flash the response rose along a curve of fixed shape that scaled linearly with flash strength¹⁵ (Fig. 2a, right panel, and Fig. 2b). Figure 2c and d show that removing Ca^{2+} scaled down the rising phases of the entire family of responses by a factor of 3.1, as if the light had been dimmed by this amount. The decrease in gain measured in this way agreed with the threefold reduction in sensitivity measured from the shift in the stimulus-response relation (Fig. 1d). The gain reduction in low $[\text{Ca}^{2+}]$ was already evident when the response rose from the baseline, about 100 ms after the flash. Analysis of the rising phase indicated that lowering $[\text{Ca}^{2+}]$ from 1 μM to 0 lowered the gain of PDE activation by factors of 1.6–10.1, with a mean of 3.7 ± 0.45 ($\pm$ s.e.m.; 22 cells).

The dependence of gain on internal $[\text{Ca}^{2+}]$ is plotted in Fig. 3. The gain decreased when $[\text{Ca}^{2+}]$ fell below about 100 nM. The curve was drawn according to the Hill equation, with half-maximal gain at 35 nM Ca^{2+} . The Hill coefficient was 1.9, indicating that Ca^{2+} reduced PDE activation cooperatively. The $[\text{Ca}^{2+}]$ in a salamander rod outer segment in darkness is estimated to be between 400 nM (ref. 6) and 150 nM (refs 16, 17). Calculations of the ability of the Na:Ca:K exchanger to pump

Ca^{2+} out of the cell⁵, as well as direct measurements using the fluorescent Ca^{2+} indicator Fura-2 (refs 16, 17), indicate that $[\text{Ca}^{2+}]$ falls below 5 nM when the light-sensitive channels close in light and $[\text{Ca}^{2+}]$ approaches equilibrium. The light-induced fall in $[\text{Ca}^{2+}]$ should therefore reduce PDE activation and by this mechanism contribute to light adaptation.

Figure 4a shows that removing Ca^{2+} desensitized the flash response without affecting recovery. The recovery phases of all responses in both 1 μM and 0 Ca^{2+} fell along a common curve when the responses were shifted appropriately on the time axis. The effect of low $[\text{Ca}^{2+}]$ was therefore to reduce PDE activation without affecting the rate of PDE deactivation, exactly as if the light were dimmed.

The lack of effect of Ca^{2+} on PDE deactivation in the present experiments is explained by the rapid loss of the Ca^{2+} -binding protein recoverin after truncation⁸. In the presence of Ca^{2+} , recoverin slows PDE deactivation by inhibiting rhodopsin phosphorylation¹¹.

Surprisingly, Ca^{2+} could affect the gain only during a brief interval near the time when light was being absorbed. A 5-s exposure to 30 nM Ca^{2+} beginning 4 s before the flash desensitized the response (Fig. 4b, trace 2), whereas the same exposure

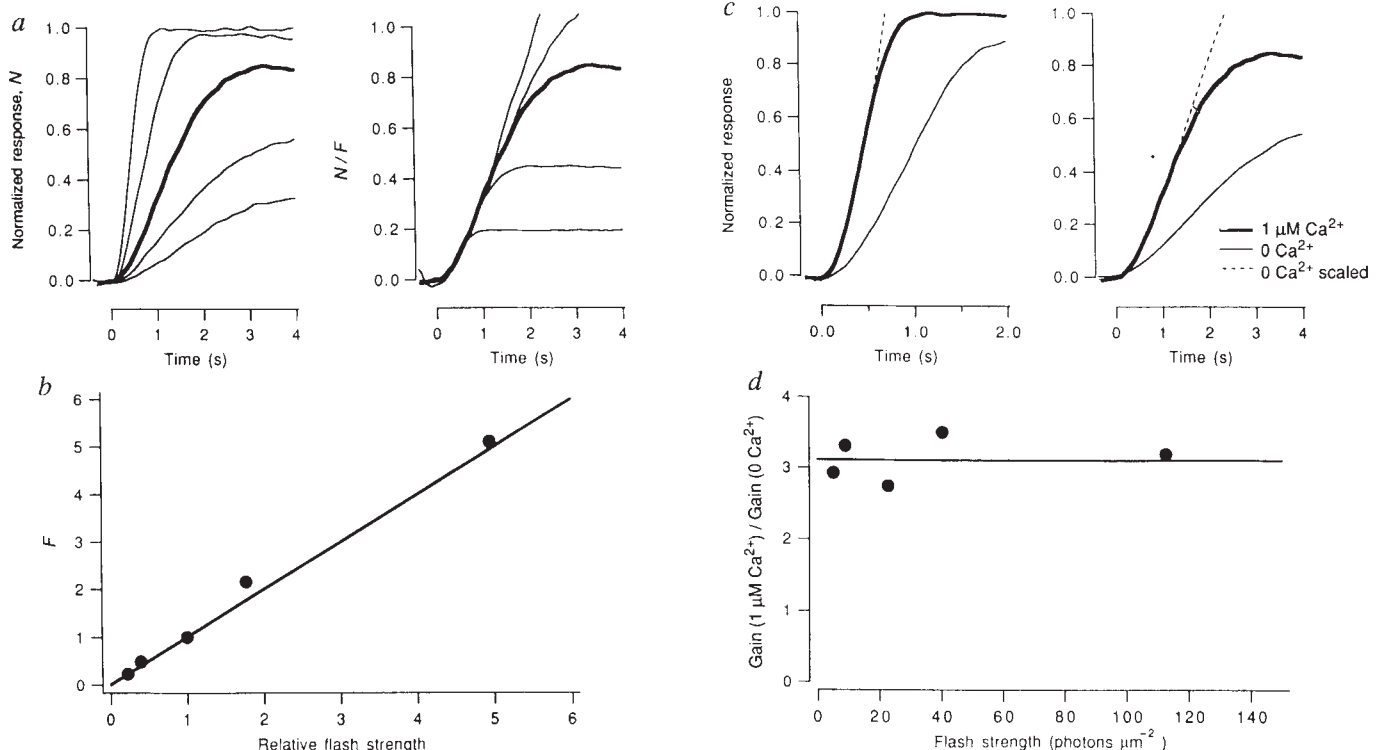


FIG. 2 Lowering $[\text{Ca}^{2+}]$ reduced PDE activation by light. **a**, Left-hand panel: Rising phases of the flash responses in 1 μM Ca^{2+} (Fig. 1c) on an expanded timescale. Flashes delivered at 0 s. The bold trace is the response to a flash of intermediate strength, delivering 23 photons μm^{-2} . Right-hand panel: Each response has been divided by a factor F that made the initial rising phase superimpose on that of the bold response. F was taken as one for the bold response. Values of F were obtained by minimizing the mean square difference between the bold response and the scaled response over the initial 0.7–1.5 s. **b**, Factors F from **a** plotted as a function of flash strength. The points fit a straight line of unit slope through the origin, showing that the initial rise of the responses varied linearly with flash strength. Similar behaviour was observed in 0 Ca^{2+} . **c**, Left-hand panel: Rising phase of responses to a flash of fixed strength (113 photons μm^{-2}) delivered in 1 μM Ca^{2+} (bold trace) or 0 Ca^{2+} (thin trace). The dashed trace shows the response in 0 Ca^{2+} multiplied by an F of 3.2, which made it superimpose on the 1 μM Ca^{2+} trace. The gain of the transduction cascade in 1 μM Ca^{2+}

was therefore 3.2 times that in 0 Ca^{2+} . Right-hand panel: Similar results for the flash delivering 23 photons μm^{-2} , when the gain in 1 μM Ca^{2+} was 2.75 times that in 0 Ca^{2+} . **d**, The gain in 1 μM Ca^{2+} , relative to that in 0 Ca^{2+} , plotted as a function of flash strength. The horizontal line through the points corresponds to a mean relative gain in this experiment of 3.1. The effect of $[\text{Ca}^{2+}]$ was independent of flash strength over this range. A trivial explanation for reduced transducin activation would be that the stimulation of guanylate cyclase on lowering Ca^{2+} lowered the concentration of GTP available to activate transducin. Evidence against this idea was provided by experiments in which 10 μM GTP was replaced by 10 μM of the phosphorothioate analogue GTP- α -S_p (ref. 28). This analogue generates active transducin with good efficiency, yet serves as a poor substrate for guanylate cyclase. It is consumed by the cyclase ~ 70 times more slowly than GTP and has the same Ca^{2+} sensitivity for cyclization²⁸. The antagonism of the light response on lowering Ca^{2+} was not affected when GTP- α -S_p replaced GTP (4 cells).

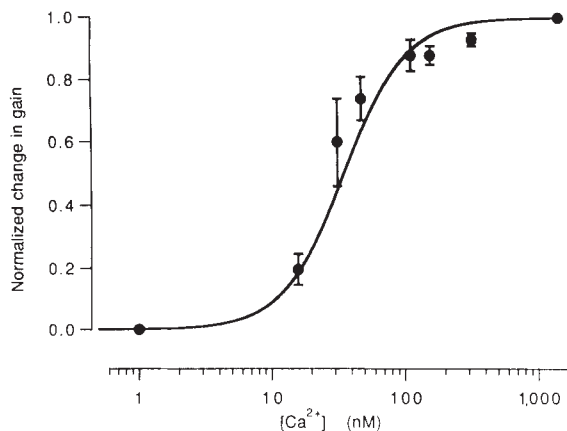


FIG. 3 The relation between free internal $[Ca^{2+}]$ and the gain of PDE activation. For each cell, the change in gain on transferring from $1 \mu M$ Ca^{2+} to a lower test $[Ca^{2+}]$ was measured by the method shown in Fig. 2c, using flashes that were just saturating. The change in gain was normalized to the maximum observed on transferring from $0 Ca^{2+}$ to $1 \mu M$ Ca^{2+} . Pooled results from 21 cells. The curve through the points was drawn according to the Hill equation, $G = [Ca^{2+}]^h / ([Ca^{2+}]^h + K^h)$, where G is the normalized change in gain, h is the Hill coefficient (1.9) and K is the $[Ca^{2+}]$ at which the change in gain was half-maximal (35 nM). Bars show one standard deviation.

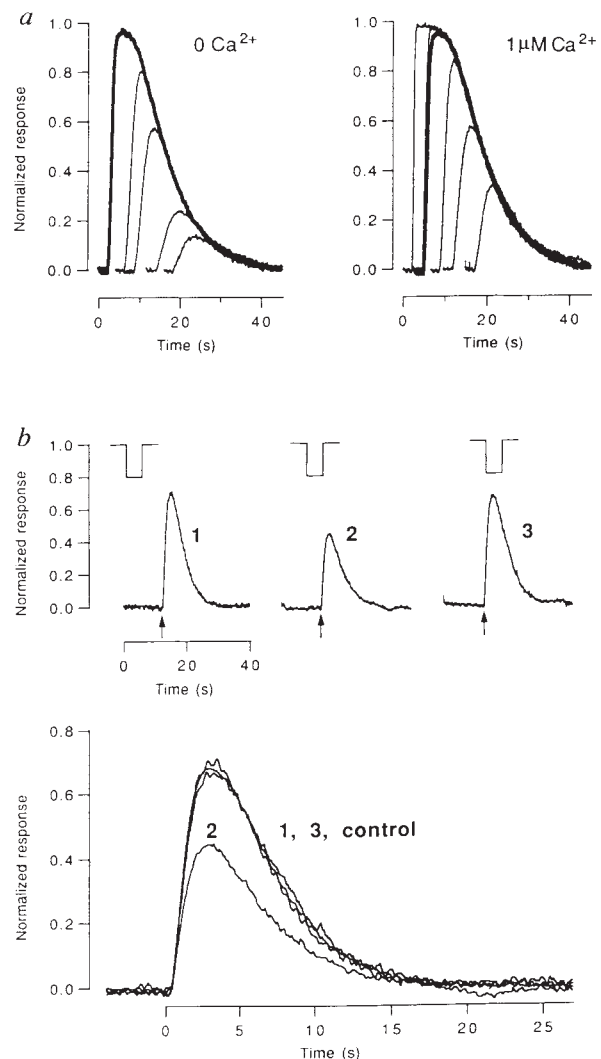
beginning 1 s after the flash had no effect (trace 3). Similar results were observed in 10 other cells. This timing suggests an action at the beginning of the cascade, in the formation of photoexcited rhodopsin (R^*). If low $[Ca^{2+}]$ had directly inhibited the hydrolytic activity of PDE or interfered with activation of PDE by transducin, the light response should have been inhibited by lowering Ca^{2+} after the flash, when both species were still active. A pulse ending 6 s before the flash had no effect (Fig. 4b, trace 1), indicating that the effect of low $[Ca^{2+}]$ required light or was rapidly reversible. An effect of Ca^{2+} at an early stage in the transduction chain has been suggested¹⁹.

The effect of Ca^{2+} on light-triggered R^* formation has high affinity and cooperativity (Fig. 3), suggesting that the Ca^{2+} -sensor is a member of the E-F hand family of Ca^{2+} -binding proteins (reviewed in ref. 20). The gradual loss of Ca^{2+} sensitivity during dialysis suggests that the protein is soluble. Two E-F hand proteins identified in rods are calmodulin¹⁰ and recoverin²¹ (also known as S-modulin¹⁸). Neither reduced the gain of PDE activation when dialysed into the outer segment in low Ca^{2+} (4 cells). Similarly, the calmodulin inhibitor calmidazolium did not affect the gain of PDE activation ($10 \mu M$, 3 cells).

Low $[Ca^{2+}]$ reduced R^* activity by a mechanism that is operationally equivalent to dimming the light (Figs 1, 2, 4a), but the time resolution of these experiments prevents us from defining the precise period of Ca^{2+} susceptibility. Low $[Ca^{2+}]$ might act before light absorption, to make a proportion of rhodopsin mol-

FIG. 4 The gain of PDE activation was determined by the free $[Ca^{2+}]$ close to the time of light absorption. **a**, Left-hand panel: Flash responses in $0 Ca^{2+}$ (from Fig. 1c) shifted on the time axis to superimpose their falling phases. The rate at which the responses recovered was independent of flash strength. The bold response recovered with a time-constant of 8.9 s. Right-hand panel: Recovery of flash responses in $1 \mu M$ Ca^{2+} also fell along a stereotyped time-course. The bold response is that in $0 Ca^{2+}$ from the left-hand panel, showing that the recovery of the flash response was unaffected by Ca^{2+} . **b**, Upper panel: The free $[Ca^{2+}]$ was reduced from $1 \mu M$ to $30 nM$ for 5 s beginning 11 s before (1), 4 s before (2) or 1 s after (3) the flash (23 photons μm^{-2} at arrow). Solution changes shown above. Traces 2 and 3 are both the average of two trials. Lower panel: Responses are superimposed and compared with a control response in which free $[Ca^{2+}]$ was $1 \mu M$ throughout (average of 3 interleaved trials).

METHODS. The records in **b** were corrected for two distortions that occurred when the $[Ca^{2+}]$ was lowered. These resulted from changes in junction current and increased affinity of the channels for cGMP¹⁰. The junction current was removed by subtracting the current recorded when $[Ca^{2+}]$ was lowered during the cell's saturating light response. The increased channel affinity caused a slow increase in the inward current which was fully reversible (see, for example, response to second low $[Ca^{2+}]$ pulse in Fig. 1b). It resulted from altered channel affinity rather than cyclase activation because: (1) no current was activated without added cGMP; (2) when the [cGMP] was varied the magnitude of the change scaled with the initial current; and (3) in the presence of cGMP, the same change occurred whether other nucleotides (ATP and GTP) were present or absent. This effect was removed by dividing the measured currents by a time-dependent scaling factor derived from the record in darkness, after junction current subtraction. The magnitude of the two distortions varied considerably between experiments. In that of **b** both were less than 10% of the amplitude of the control response. Slower changes in the affinity of the channel for cGMP have also been described²⁹, but were not observed in these experiments.



ecules refractory. If this is so, rhodopsin escapes the refractory state within seconds of return to high $[Ca^{2+}]$ (Fig. 4b). Alternatively, low $[Ca^{2+}]$ might act after light absorption to deactivate R^* . If this is so, R^* must be only transiently susceptible to Ca^{2+} , as the gain reduction was complete within 100 ms (Fig. 2c) and lowering Ca^{2+} after the flash had no effect (Fig. 4b).

It will be interesting to determine the molecular mechanism of this new effect and to see whether Ca^{2+} similarly regulates the activity of seven-helix receptors in other G-protein transduction systems. □

Received 26 July; accepted 1 November 1993.

1. Matthews, H. R., Murphy, R. L., Fain, G. L. & Lamb, T. D. *Nature* **334**, 67–69 (1988).
2. Nakatani, K. & Yau, K.-W. *Nature* **334**, 69–71 (1988).
3. Yau, K.-W. & Nakatani, K. *Nature* **313**, 579–582 (1985).
4. McNaughton, P. A., Cervetto, L. & Nunn, B. J. *Nature* **322**, 261–263 (1986).
5. Cervetto, L., Lagnado, L., Perry, R. J., Robinson, D. W. & McNaughton, P. A. *Nature* **337**, 740–743 (1989).
6. Lagnado, L., Cervetto, L. & McNaughton, P. A. *J. Physiol., Lond.* **455**, 111–142 (1992).
7. Lagnado, L. & Baylor, D. A. *Neuron* **8**, 995–1002 (1992).
8. Lolley, R. N. & Racz, E. *Vision Res.* **22**, 1481–1486 (1982).

9. Koch, K. W. & Stryer, L. *Nature* **334**, 64–66 (1988).
10. Hsu, Y.-T. & Molday, R. S. *Nature* **361**, 76–79 (1993).
11. Kawamura, S. *Nature* **362**, 855–857 (1993).
12. Lamb, T. D., McNaughton, P. A. & Yau, K.-W. *J. Physiol., Lond.* **319**, 463–496 (1981).
13. Fain, G. L. *J. Physiol., Lond.* **261**, 71–101 (1976).
14. Lamb, T. D. & Pugh, E. N. *J. Physiol., Lond.* **449**, 719–757 (1992).
15. Baylor, D. A. & Hodgkin, A. L. *J. Physiol., Lond.* **242**, 729–758 (1974).
16. McCarthy, S. T. thesis, Univ. California, Berkeley (1993).
17. Younger, J. P., McCarthy, S. T. & Owen, W. G. *Invest. Ophthalm. Vis. Sci.* **33**, 1104 (1992).
18. Kawamura, S. & Murakami, M. *Nature* **349**, 420–423 (1991).
19. Hodgkin, A. L., McNaughton, P. A. & Nunn, B. J. *J. Physiol., Lond.* **372**, 54P (1986).
20. Persechini, A., Moncrief, N. D. & Kretsinger, R. H. *Trends Neurosci.* **12**, 462–467 (1989).
21. Dizhoor, A. M. et al. *Science* **251**, 915–918 (1991).
22. Nakatani, K. & Yau, K.-W. *J. Physiol., Lond.* **395**, 731–753 (1988).
23. Baylor, D. A., Lamb, T. D. & Yau, K.-W. *J. Physiol., Lond.* **288**, 589–611 (1979).
24. Hodgkin, A. L., McNaughton, P. A. & Nunn, B. J. *J. Physiol., Lond.* **358**, 447–468 (1985).
25. Fabiato, A. & Fabiato, J. *J. Physiol., Paris* **75**, 463–505 (1979).
26. Wilden, U., Hall, S. W. & Kuhn, H. *Proc. natn. Acad. Sci. U.S.A.* **83**, 1174–1178 (1986).
27. Lamb, T. D., Matthews, H. R. & Torre, V. J. *J. Physiol., Lond.* **372**, 315–349 (1986).
28. Koch, K.-W., Eckstein, F. & Stryer, L. *J. biol. Chem.* **265**, 9659–9663 (1990).
29. Gordon, S. E., Brautigan, D. L. & Zimmerman, A. L. *Neuron* **9**, 739–748 (1992).

ACKNOWLEDGEMENTS. This work was supported by grants from the National Eye Institute and the Alcon Research Institute. L.L. held a Human Frontiers in Science Program Long-Term Fellowship.

Requirement for the orphan steroid receptor Nur77 in apoptosis of T-cell hybridomas

John D. Woronicz, Barbara Calnan, Vu Ngo & Astar Winoto*

Department of Molecular and Cell Biology,
Division of Immunology and Cancer Research Laboratory,
University of California, Berkeley, California 94720, USA

APOPTOSIS is a phenomenon observed during development of many cell types in many organisms. It is an internal, programmed cell death characterized by DNA fragmentation into nucleosome-size pieces^{1–3}. Anti-CD3-induced apoptosis in T-cell hybridomas and immature thymocytes requires new gene transcription and may be related to negative selection during T-cell development^{4–6}. Using subtractive hybridization, we isolated a complementary DNA clone encoding the orphan steroid receptor Nur77 (refs 7–9). It shows different patterns of messenger RNA induction between apoptotic and stimulated T cells. We report here the use of gel shift analysis to demonstrate that the Nur77 protein is present at high levels in apoptotic T-cell hybridomas and apoptotic thymocytes, but not in growing T cells or stimulated splenocytes. A Nur77 dominant negative protected T-cell hybridomas from activation-induced apoptosis. Hence Nur77 is necessary for induced apoptosis in T-cell hybridomas.

To isolate genes involved in cell death, we used the T-cell hybridomas AO4H5.3, DO11.10 and KZH (ref. 10). Apoptotic DNA ladders are observed as early as 4 hours post-stimulation with anti-CD3 antibodies, which mimics signalling through the T-cell receptor. Most cells show apoptotic morphology and death by 18 h post-stimulation. The anti-CD3 antibody-induced apoptosis occurs in T-cell hybridomas and thymocytes but not in splenocytes or a T-cell lymphoma EL4, or when T-cell hybridomas are induced with phorbol ester (Fig. 1a). We constructed a cDNA library from apoptotic AO4H5.3. Using a subtractive probe (apoptotic-resting AO4H5.3), one clone was isolated (we initially called it 'ced'). Sequencing revealed that it was a previously identified orphan steroid receptor (Nur77, N10, NGF-IB) with no known function^{9,11,12}.

To correlate Nur77 gene expression with activation-induced apoptosis, we examined the kinetics of Nur77 mRNA induction

after stimulation in AO4H5.3 and EL4. In AO4H5.3 stimulated with phorbol-12-myristate 13-acetate (PMA), or anti-CD3-stimulated EL4, the Nur77 mRNA is induced starting at 30 min post-stimulation, peaking at 1 hour and then rapidly downregulated (Fig. 1b, and data not shown). This resembles the induction kinetics in fibroblasts⁹. In contrast, the Nur77 mRNA levels remain high in anti-CD3-stimulated AO4H5.3 to at least 14 hours post-induction (Fig. 1b). Other early-response genes tested (*c-jun*, *c-myc*, *egr-1*, *HLH462*^{7,13,16}) showed no discernible differences in their expression between apoptotic and activated T cells (J.D.W., unpublished data). Induction of Nur77 mRNA is specific for activation-induced cell death because dexamethasone-treated AO4H5.3 or withdrawal of interleukin-2 (IL-2) from an IL-2-dependent cell line CTLL-2 did not induce Nur77 gene expression.

To characterize Nur77 protein activity, we used nuclear extracts from several T-cell hybridomas, EL4, thymocytes and splenocytes in a gel shift analysis using an oligonucleotide with an Nur77 DNA-binding site^{7,18}. A striking difference was seen between dying and growing T cells. In anti-CD3-stimulated EL4, anti-CD3-stimulated splenocytes or PMA-stimulated AO4H5.3, hardly any Nur77–DNA binding activity could be detected (Fig. 1c–e). In contrast, Nur77–DNA complexes were detected in apoptotic AO4H5.3 starting at 2 h post-stimulation and continued to accumulate through 12 h post-stimulation (Fig. 1c). The Nur77–DNA complexes could also be seen in the apoptotic hybridomas DO11.10, KZH and in apoptotic thymocytes (Fig. 1d). Binding activity was specific, as competition using an unlabelled Nur77 oligonucleotide abolished binding activity, but addition of an irrelevant oligonucleotide did not affect the Nur77 activity (Fig. 1c, and see below). As a control for the integrity of the nuclear extracts, we also used the GT box oligonucleotide, which detects the ubiquitous Sp1 family proteins¹⁹ (Fig. 1f, and data not shown). Thus, Nur77 DNA-binding activity is induced at a high level during induced apoptosis.

To study the role of Nur77 in apoptosis, we created a dominant negative mutant to suppress the endogenous Nur77 activity. Nur77, when expressed from the cytomegalovirus (CMV) promoter (pced-1 made in pcDNA-1; Invitrogen), can transactivate a reporter construct containing its DNA-binding site. The reporter constructs used contain either 2, 3 or 4 Nur77 DNA-binding sites upstream of a minimal promoter and the CAT (chloramphenicol acetyltransferase) reporter gene ($\Delta 56$ -ced², $\Delta 56$ -ced³ or $\Delta 56$ -ced⁴). Transfection of pced-1 with the $\Delta 56$ -ced reporter constructs resulted in a marked increase of the CAT activity, with $\Delta 56$ -ced⁴ having the highest transactivation activity (Fig. 2a). As expected, transfection of the vector alone (pcDNA-

* To whom correspondence should be addressed.

1) did not activate transcription. To create a dominant negative, a Nur77 mutant containing only its DNA-binding domain (pEV-cedZnCT) or only its N-terminal activation domain (pced- Δ XR) were created. The former might function by binding but not transactivating from a Nur77 DNA-binding site, the latter might sequester co-activator proteins and indirectly decrease the wild-type Nur77 activity. As expected, neither protein transactivated Δ 56-ced⁴ (Fig. 2a: 6 and 7). To test for the dominant negative activity, we did triple transfection experiments with these deletion constructs, Δ 56-ced⁴ reporter plasmid and pced-1 wild-type Nur77 expression construct. Transfection of pced- Δ XR had no effect on Nur77 transactivation (8 in Fig. 2a). In contrast, cotransfection of pEV-cedZnCT construct decreased Nur77

transactivation to almost background levels (9 in Fig. 2a) and could decrease Nur77 transactivation even when transfected at one eighth the amount of Nur77 (Fig. 2b). Hence, the Nur77 lacking its N-terminal transactivation domain (ZnCT) is a powerful dominant negative mutant.

The ZnCT dominant negative was transferred into a neomycin expression vector (BMGNeo²⁰) and stably transfected into AO4H5.3. AO4H5.3 transfected with the BMGNeo vector alone was used as a control. Two BMGNeo clones (Neo-F12 and Neo-B11) and three BMGNeo-cedZnCT dominant negative clones (ZnCT-A5, ZnCT-B8, ZnCT-G12) were obtained. Gel shift analysis using the Nur77 oligonucleotide showed, as expected, a major Nur77 protein/DNA complex induced by CD3 antibody

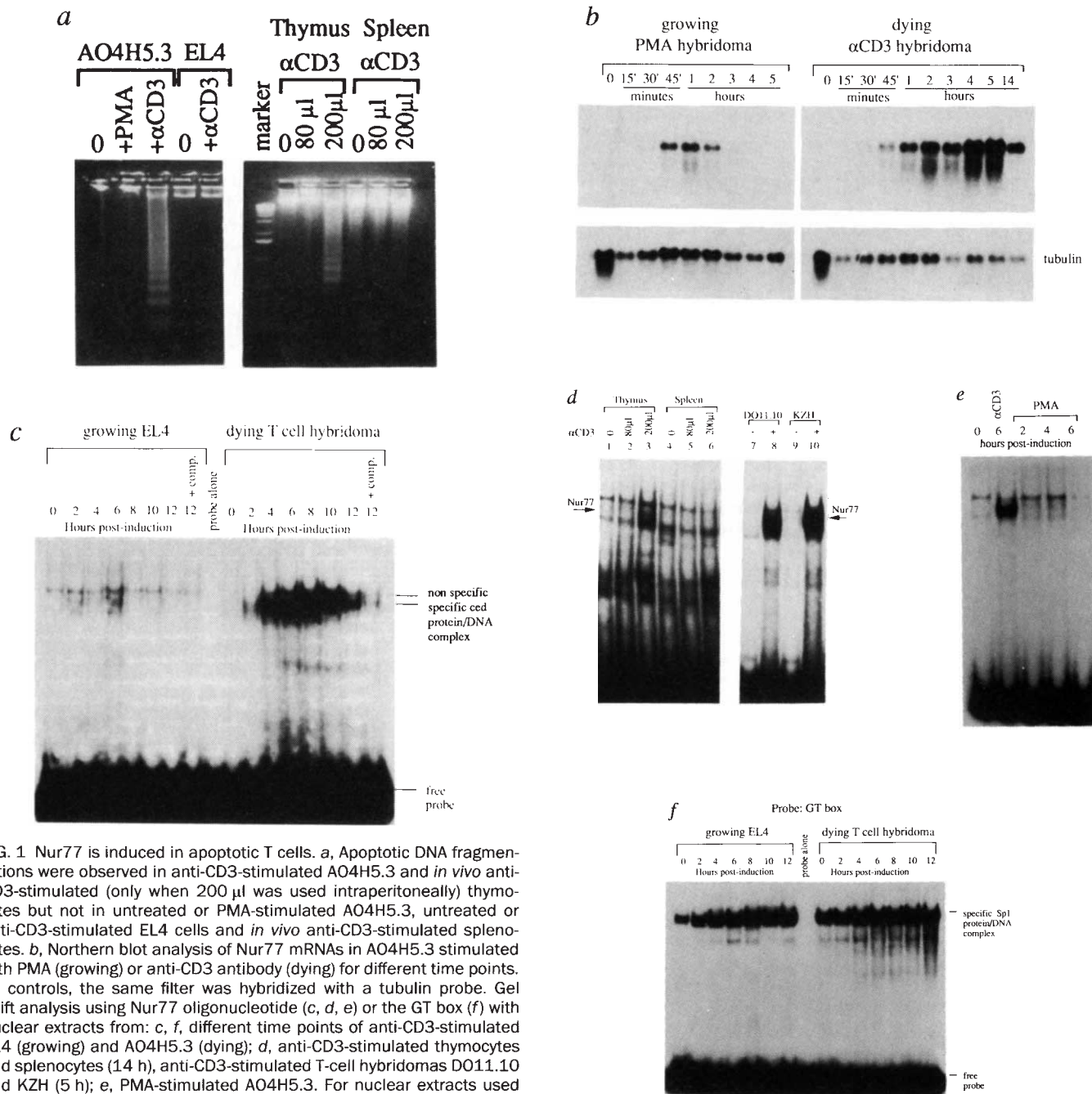
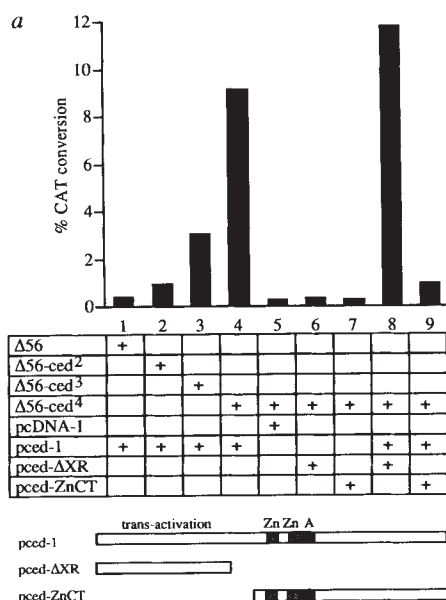


FIG. 1 Nur77 is induced in apoptotic T cells. **a**, Apoptotic DNA fragmentations were observed in anti-CD3-stimulated AO4H5.3 and *in vivo* anti-CD3-stimulated (only when 200 μ l was used intraperitoneally) thymocytes but not in untreated or PMA-stimulated AO4H5.3, untreated or anti-CD3-stimulated EL4 cells and *in vivo* anti-CD3-stimulated splenocytes. **b**, Northern blot analysis of Nur77 mRNAs in AO4H5.3 stimulated with PMA (growing) or anti-CD3 antibody (dying) for different time points. As controls, the same filter was hybridized with a tubulin probe. Gel shift analysis using Nur77 oligonucleotide (**c**, **d**, **e**) or the GT box (**f**) with nuclear extracts from: **c**, **f**, different time points of anti-CD3-stimulated EL4 (growing) and AO4H5.3 (dying); **d**, anti-CD3-stimulated thymocytes and splenocytes (14 h), anti-CD3-stimulated T-cell hybridomas DO11.10 and KZH (5 h); **e**, PMA-stimulated AO4H5.3. For nuclear extracts used in **d**, control gel shift analysis with a GATA-3 probe³¹ showed that they contain equal amounts of protein.

METHODS. DNA fragmentation, gel shift and northern blot analysis were done as described^{19,29}. The oligonucleotide-containing Nur77 DNA-binding site was synthesized according to ref. 17 with *Sall* and *XhoI*

restriction sites at either end. The oligonucleotide containing GT box is from ref. 19. The cell surface expression of the CD3 molecule for both T-cell hybridomas and EL4 lymphoma was checked regularly using FACS analysis.



in all of the clones (Fig. 3, lanes 1–10). In addition, a lower complex corresponding to the expected truncated Nur77 protein is seen in all the ZnCT clones (A5, B8 and G12) with and without anti-CD3 stimulation. The specificity of the protein complexes was confirmed by competition with an excess unlabelled Nur77 oligonucleotide, but not with an irrelevant unlabelled GT box oligonucleotide (Fig. 3, lanes 11 and 12). To see whether ZnCT could inhibit or slow the activation-induced cell death, we used MTT (Fig. 4 legend), trypan blue and DNA fragmentation assays to quantify the rate of apoptosis in the transfected cells. MTT is metabolized to a coloured precipitate in living but not dead cells and can be quantified by optical density. The MTT assay was done 19 h after anti-CD3 stimulation. The two Neo control clones B11 and F12 die after anti-CD3 stimulation, and the optical density at 550 nm dropped from 2 to 0.4 or 1.8 to 0.7. In contrast, the ZnCT clones, A5, G12 and B8 continued to metabolize MTT, indicating viability for most of the cells

FIG. 3 Gel shift analysis with Nur77 oligonucleotide and nuclear extracts from unstimulated and 5-h anti-CD3-stimulated (– or +, respectively) AO4H5.3 clones either stably transfected with BMGNeo alone (Neo-F12 and Neo-B11) or a Nur77 dominant negative mutant (ZnCT-A5, ZnCT-G12, ZnCT-B8). The construct contains a blunt-ended 1.4-kb *SacI* fragment of pced-ZnCT in the blunt *SacI* site of BMGNeo.

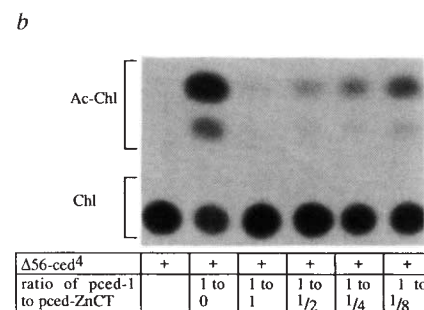
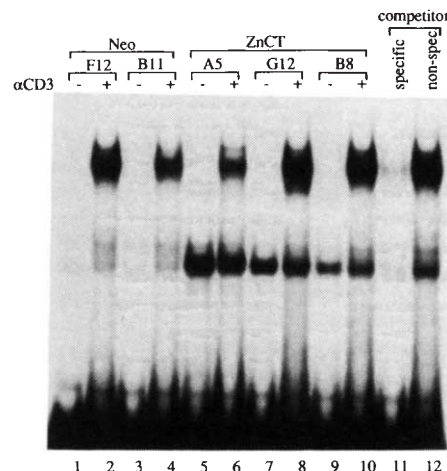


FIG. 2 **a, b**, Nur77-dominant negative mutant assayed in transient transfections. The plasmids indicated as + were transfected into 293⁺ cells and CAT activities were measured 2 days later. The same molar ratio and weight of DNA were used (for **b**, the molar ratios of wild-type Nur77 to ZnCT plasmids were varied as indicated; sheared calf thymus DNA was used to equalize the total DNA). Plasmid β-actin-β-gal was usually included to standardize the transfection efficiency. The Δ56 reporter construct contains a minimal *c-fos* promoter (–56 to +109). Multimers (dimer, 3-mer or 4-mers) of Nur77 oligonucleotides were cloned into the *SacI* site of the Δ56 plasmid to create Δ56-ced², Δ56-ced³ and Δ56-ced⁴. For expression plasmids, either pcDNA-1 (Invitrogen) or pEVRF (ref. 30) plasmids were used. pced-1 is the pcDNA-1 plasmid containing a full-length Nur77 cDNA, whereas pced-ΔXR contains only the 5' portion of Nur77 cDNA up to the *XmnI* site. For the dominant negative mutant, pced-ZnCT, the 3' Nur77 cDNA (from *HgaI* site upstream of the Zn-finger to the end) was cloned into the pEVRF vector in frame with the AUG initiation codon. Zn and A indicate zinc-finger and A box DNA-binding domains¹⁸. Ac-Chl and Chl denote acetylated and unconverted chloramphenicol.

(Fig. 4a). These results were confirmed by trypan blue, which stains dead cells. Most cells from all three ZnCT clones remained viable (55–70%) after 18 hours post-stimulation, whereas the Neo controls showed less than 10% cell viability (Fig. 4b). No apoptotic DNA ladders were seen at 4 and 5 h anti-CD3 stimulation for the three ZnCT clones. In contrast, the parental hybridomas AO4H5.3 and the two Neo control clones B11 and F12 showed extensive DNA fragmentation at similar time points (Fig. 4c). The activation pathway for AO4H5.3 was not affected in these clones, because stimulation with anti-CD3 still produced high levels of IL-2 (data not shown). Similar transfections into DO11.10 showed essentially the same results. No apoptotic DNA ladders can be seen in the two dominant negative clones at 5, 6 and 7 h post-stimulation (Fig. 4d). Hence transcriptional activation of the Nur77 gene product is required for activation-induced cell death.

We have shown that Nur77, an orphan steroid receptor gene



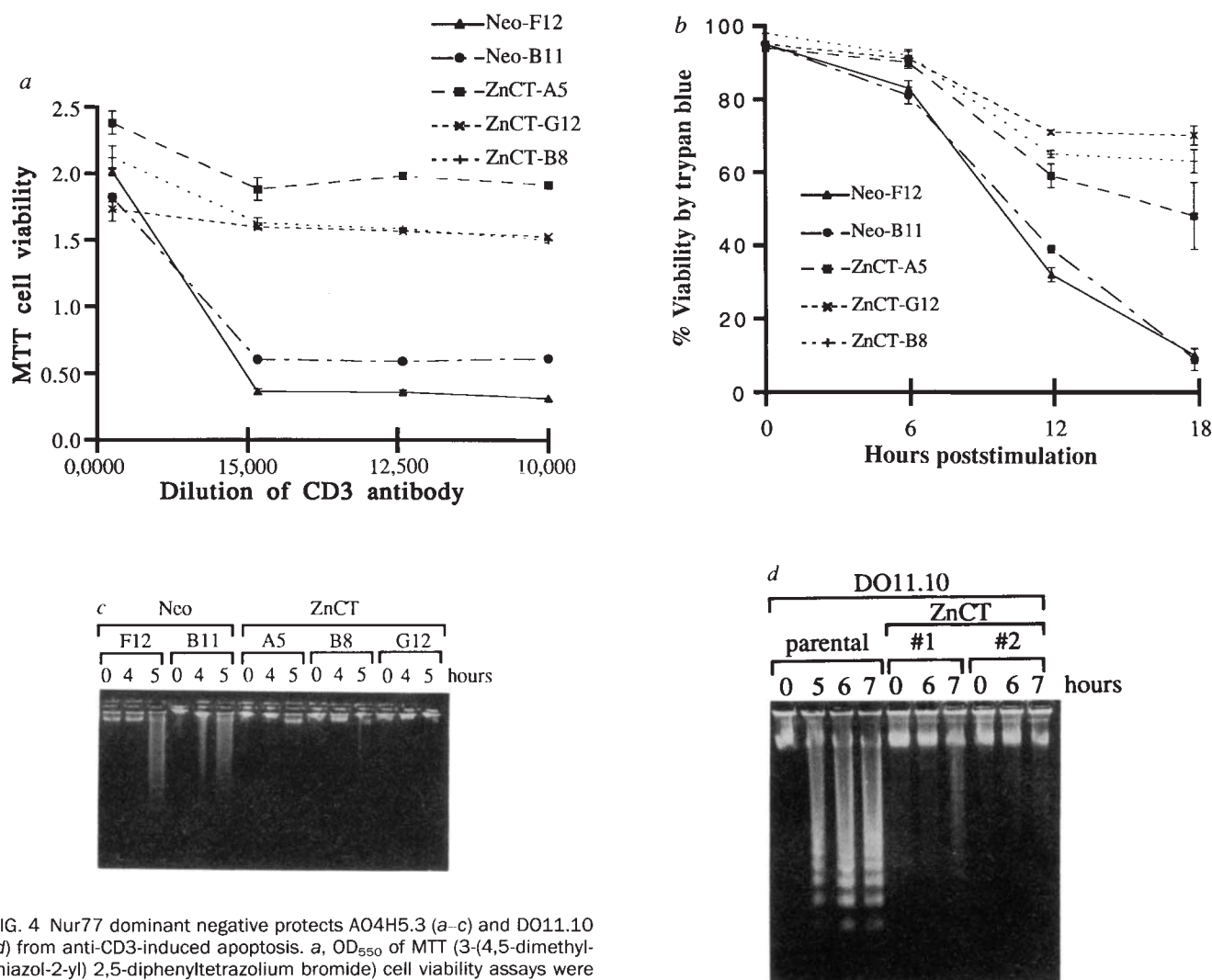


FIG. 4 Nur77 dominant negative protects A04H5.3 (a-c) and DO11.10 (d) from anti-CD3-induced apoptosis. **a**, OD₅₅₀ of MTT (3-(4,5-dimethylthiazol-2-yl) 2,5-diphenyltetrazolium bromide) cell viability assays were determined after 19 h stimulation with CD3 antibody-coated plates. **b**, Trypan blue cell viability was tested at the indicated time points. **c**, DNA fragmentation after stimulation with anti-CD3-coated plates for the indicated times. **d**, DNA ladder assays were done with DO11.10 parental and the two Nur77 dominant negative ZnCT clones after stimulation with anti-CD3-coated plates for the indicated time points. METHODS. MTT and DNA fragmentation assays were done as described^{25,29}. Values are averaged from triplicate wells and are repre-

sentative of several independent experiments. Stable transfections into DO11.10 were done using electroporation using the BMG-Neo ZnCT plasmid. G418 resistant clones were checked for the presence of ZnCT activity using gel shift analysis. The cell surface CD3 expression in these clones was checked using fluorescence activated cell sorting and the ability to induce the endogenous Nur77 activity.

product, is required for T-cell-receptor-mediated apoptosis. Signalling mediated by anti-CD3 differentially induces Nur77 expression in apoptotic T-cell hybridomas and thymocytes but not in growing T cells or stimulated splenocytes. As apoptosis occurs during negative selection of T cells, Nur77 may play a key role during T-cell development. This is being examined using ZnCT in transgenic mice. Nur77 is not required for glucocorticoid-induced cell death. Addition of dexamethasone to the ZnCT clones can still induce apoptosis (data not shown), indicating separate pathways between anti-CD3 and glucocorticoid-induced cell death. But addition of dexamethasone to the anti-CD3-treated T-cell hybridomas inhibits apoptosis, and at the same time severely decreases the Nur77 protein activity as measured by gel shift analysis (J.D.W. and V.N., unpublished observation). It is unknown, however, whether Nur77 is also required in other apoptotic systems. Recently, several genes implicated in apoptosis of different cell types have been identified, including *p53* (refs 21–23), *c-myc* (refs 24, 25) and *bcl-2* (ref. 26) as well as membrane proteins PD-1 (ref. 27) and Fas antigen²⁸. It will be interesting to determine the functional rela-

tionship between Nur77 and these gene products during apoptosis. □

Received 5 April; accepted 1 October 1993.

1. Cohen, J. J., Duke, R. C., Fadok, V. A. & Sellins, K. S. *A. Rev. Immun.* **10**, 267–293 (1992).
2. Oppenheim, R. W. *A. Rev. Neurosci.* **14**, 453–501 (1991).
3. Kerr, J. F. R. & Harmon, B. V. in *Apoptosis: the Molecular Basis of Cell Death* (eds Tomei, L. D. & Cope, F. O.) 5–29 (Cold Spring Harbor Laboratory Press, New York, 1991).
4. Mercep, M., Noguchi, P. D. & Ashwell, J. D. *J. Immun.* **142**, 4085–4092 (1989).
5. Shi, Y. F. et al. *J. Immun.* **144**, 3326–3333 (1990).
6. Smith, C. A. et al. *Nature* **337**, 181–184 (1989).
7. Christy, B. A., Lau, L. F. & Nathans, D. *Proc. natn. Acad. Sci. U.S.A.* **85**, 7857–7861 (1988).
8. Milbrandt, J. *Neuron* **1**, 183–188 (1988).
9. Ryseck, R. P. et al. *EMBO J.* **8**, 3327–3335 (1989).
10. Shastri, N., Gammon, G., Miller, A. & Sercarz, E. E. *J. exp. Med.* **164**, 882–896 (1986).
11. Hazel, T. G., Nathans, D. & Lau, L. F. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8444–8448 (1988).
12. Watson, M. A. & Milbrandt, J. *Molec. cell. Biol.* **9**, 4213–4219 (1989).
13. Milbrandt, J. *Science* **238**, 797–799 (1987).
14. Sukhatme, V. P. et al. *Cell* **53**, 37–43 (1988).
15. Lemaire, P., Revelant, O., Bravo, R. & Charnay, P. *Proc. natn. Acad. Sci. U.S.A.* **85**, 4691–4695 (1988).
16. Christy, B. A. et al. *Proc. natn. Acad. Sci. U.S.A.* **88**, 1815–1819 (1991).
17. Wilson, T. E., Fahrner, T. J., Johnston, M. & Milbrandt, J. *Science* **252**, 1296–1300 (1991).
18. Wilson, T. E., Paulsen, R. E., Padgett, K. A. & Milbrandt, J. *Science* **256**, 107–110 (1992).
19. Kingsley, C. & Winoto, A. *Molec. cell. Biol.* **12**, 4251–4261 (1992).
20. Karasuyama, H. & Melchers, F. *Eur. J. Immun.* **18**, 97–104 (1988).

21. Shaw, P. et al. *Proc. natn. Acad. Sci. U.S.A.* **89**, 4495–4499 (1992).
22. Yonish, R. E. et al. *Nature* **353**, 345–347 (1991).
23. Lowe, S. W. et al. *Nature* **362**, 847–849 (1993).
24. Evan, G. I. et al. *Cell* **69**, 119–128 (1992).
25. Shi, Y. et al. *Science* **257**, 212–214 (1992).
26. Korsmeyer, S. J. *Immun. Today* **13**, 285–288 (1992).
27. Ishida, Y., Agata, Y., Shibahara, K. & Honjo, T. *EMBO J.* **11**, 3887–3895 (1992).
28. Cohen, P. L. & Eisenberg, R. A. *Immun. Today* **13**, 427–428 (1992).
29. Sentman, C. L. et al. *Cell* **67**, 879–888 (1991).
30. Matthias, P. et al. *Nucleic Acids Res.* **17**, 6418 (1989).
31. Marine, J. & Winoto, A. *Proc. natn. Acad. Sci. U.S.A.* **88**, 7284–7288 (1991).

ACKNOWLEDGEMENTS. We thank G. Smith, N. Shastri and M. Shin for critical reading of the manuscript, N. Shastri for T-cell hybridomas, R. Bravo and L. F. Lau for the full-length *Nur77* cDNA clone, H. Karasuyama for BMGNeo vector and G. Nolan for the 293⁺ cell line. A. W. is a Searle Scholar and Cancer Research Institute Investigator.

Apoptotic signals delivered through the T-cell receptor of a T-cell hybrid require the immediate-early gene *nur77*

Zheng-Gang Liu*, Sallie W. Smith†, Kelly A. McLaughlin*, Lawrence M. Schwartz*‡ & Barbara A. Osborne*†§

Departments of † Veterinary and Animal Sciences and ‡ Biology and * Program in Molecular and Cellular Biology, University of Massachusetts, Amherst, Massachusetts 01003, USA

ENGAGEMENT of the T-cell antigen receptor (TCR) on immature thymic T cells induces death by apoptosis^{1–3}. Although several lines of evidence indicate that apoptosis requires *de novo* gene expression, little is known about the molecular pathways that mediate this response. Here we show that *nur77* (refs 4–7), a zinc-finger transcription factor, is expressed in response to TCR engagement in immature T cells and T-cell hybrids. Antisense inhibition of *nur77* expression prevents apoptosis in TCR-stimulated cells. *nur77* is also expressed in response to mitogens, but in this case transcription is regulated by 5' upstream elements that are distinct from those used for induction of apoptosis. In addition, polyadenylation is only observed on *nur77* transcripts found in condemned cells. These data support a role for *nur77* in cell death that may be distinct from that of activation.

Negative selection, the clonal deletion of self-reactive T cells, occurs by apoptosis^{1–3}. Under normal circumstances, this happens when an individual CD4⁺CD8⁺ thymocyte encounters antigen presented by self-major histocompatibility complex (MHC); this death requires *de novo* gene expression⁸. To isolate genes mediating apoptosis in the thymus, we used DO11.10 transgenic mice, which provide large numbers of dying thymocytes⁹. The DO11.10 mouse expresses a T-cell receptor specific for OVA323–339, a peptide derived from chicken ovalbumin. Intraperitoneal injection with OVA323–339 results in massive apoptosis in the thymus within 6–8 h of injection⁹. Messenger RNA from the thymocytes of treated mice was used to construct a complementary DNA library which was screened by a differential hybridization strategy¹⁰ designed to isolate genes that appeared during cell death but were absent in the normal thymus.

The collection of apoptosis-associated genes isolated included a gene we designated *apt-2* (for apoptosis-2). Determination of the DNA sequence of this gene revealed that it shared 100% identity with *nur77* (data not shown), a gene previously isolated from a cDNA library constructed with the mRNA from growth-factor-induced 3T3 cells⁴. *nur77*, also known as NGFI-B (ref. 5), TIS1 (ref. 6) and N10 (ref. 7), is a member of the steroid/thyroid hormone receptor superfamily which encodes a number

TABLE 1 Correlation of expression of *nur77* and cell death

Treatment	Expression of <i>nur77</i>			Cell death (%) [*]
	45 min	4 h	23 h	
DO11.1 untreated	–	–	–	2.5
DO11.1 + F23.1	++	++	–	56
DO11.1 + PMA	+	+++	+	3.8
DO11.1 + A23187	+++	++	+	4.8
DO11.1 + PMA + A23187	+++	++	+	41

nur77 expression was determined by densitometry of autoradiograms of northern blots. Methods were as described for Fig. 1. The ratio of *nur77* to the constitutively expressed gene glyceraldehyde phosphate dehydrogenase (GAPDH) was determined. Ratios >1 are assigned +++; 0.1–1, ++; and 0.01–0.1, +.

^{*} Per cent cell death was assayed by the FITC assay²⁰.

of ligand-dependent transcription factors that are characterized by a centrally located highly conserved DNA-binding domain containing two zinc-fingers. When the thymic mRNA used to construct the library was analysed by northern blotting, *nur77* transcripts were abundant (Fig. 1a). For many of our experiments, we took advantage of DO11.10, a T-cell hybrid¹¹ that expresses the same TCR genes as the transgenic mice used in our library construction. When DO11.10 is treated with an antibody directed against the TCR (F23.1), these cells die by apoptosis¹². Correlated with this process is the induction of *nur77* mRNA (Fig. 1b). This transcript is observed when DO11.10 cells are induced to die by engagement of the TCR either by α -TCR (F23.1) or α -CD3 (245 2C11) antibodies (Fig. 1b). Expression of *nur77* is dependent upon transcription, as demonstrated by treatment with the transcriptional inhibitor actinomycin D (Fig. 1b), which correlates with data demonstrating that actinomycin

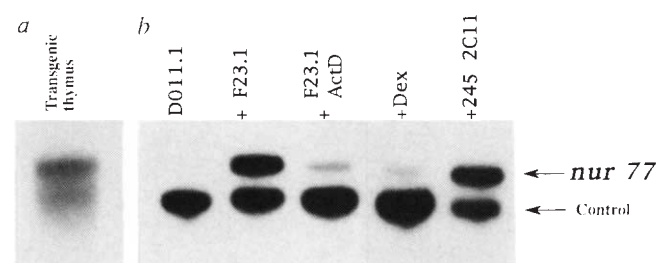


FIG. 1 Expression of *nur77* following various treatments. a, RNA isolated from thymocytes of DO11.10 TCR transgenic mice after *in vivo* induction of apoptosis by ovalbumin peptide. b, RNA isolated from the T-cell hybrid DO11.10 after treatment with the anti-TCR antibody F23.1 (ref. 9), F23.1 plus the transcriptional inhibitor actinomycin D (ActD) dexamethasone (Dex), or anti-CD3 antibody 245 2C11 (ref. 21). METHODS. Nine mice transgenic for the DO11.10 TCR (ref. 9) were injected intraperitoneally (i.p.) with 1 mM ovalbumin OVA 323–339 peptide (1 ml). Three mice were killed at 3, 4 and 5 h after the injection and their thymocytes isolated. mRNA was prepared from the pooled thymocytes and a cDNA library constructed in the UniZap vector (Stratagene); 2 μ g of this pooled mRNA was used for northern analysis (a). In separate experiments, flasks of surface area 25 cm² were coated with 2 ml 10 μ g ml⁻¹ sheep anti-mouse immunoglobulin for 1 h, washed 3 times with PBS and coated with 5 ml of culture supernatant from F23.1 (anti-TCR) or 5 ml 5 μ g ml⁻¹ 245 2C11 (anti-CD3). Cells were added to these flasks and incubated for 1 h. Alternatively, DO11.10 cells were incubated for 1 h in flasks with 62 μ M dexamethasone. In all experiments, cells were cultured in DME/RPMI 1640 with 10% horse serum and 20 μ g ml⁻¹ gentamicin and collected after 1 h incubation; total RNA was prepared using RNazol B (Cinna/Biotech Laboratories) and fractionated on a 1% agarose/formaldehyde gel, blotted and hybridized to ³²P-labelled *nur77*. GAPDH was used as a control in a, and C1, an uncharacterized constitutively expressed cDNA isolated from our thymus cDNA library, was used for b.

§ To whom correspondence should be addressed.

D retards the onset of cell death⁸. Dexamethasone, a glucocorticoid that induces cell death in lymphoid cells¹³, does not substantially induce *nur77*.

As *nur77* is induced by mitogenic stimuli in many cells, we wanted to determine whether its expression correlated with death or inappropriate activation. DO11.10 cells were treated with different agents that either induce death or activate T cells, and RNA was isolated at various times after treatment. All treatments that induced death also induced expression of *nur77* (Table 1). We also noted, however, that *nur77* was expressed under some conditions that did not lead to death, for example, treatment with either the phorbol ester PMA, or the Ca^{2+} iono-

phore A23187. To discriminate between *nur77* expression induced by activation as opposed to death, we examined the 5' upstream region to determine whether there were independent regulatory elements. A 3,500-base-pair (bp) fragment derived from the 5' promoter of *nur77* was cloned upstream of the chloramphenicol acetyltransferase (CAT) gene containing the 3' polyadenylation site from SV40 (Fig. 2a). This construct, and nested deletions made from the 5' end of the *nur77* promoter, were used to transfect DO11.10 cells transiently. Twenty-four hours after transfection, cells were stimulated with F23.1, the anti-TCR antibody or with PMA, and CAT activity was determined. The data obtained (Fig. 2b) show that induction in

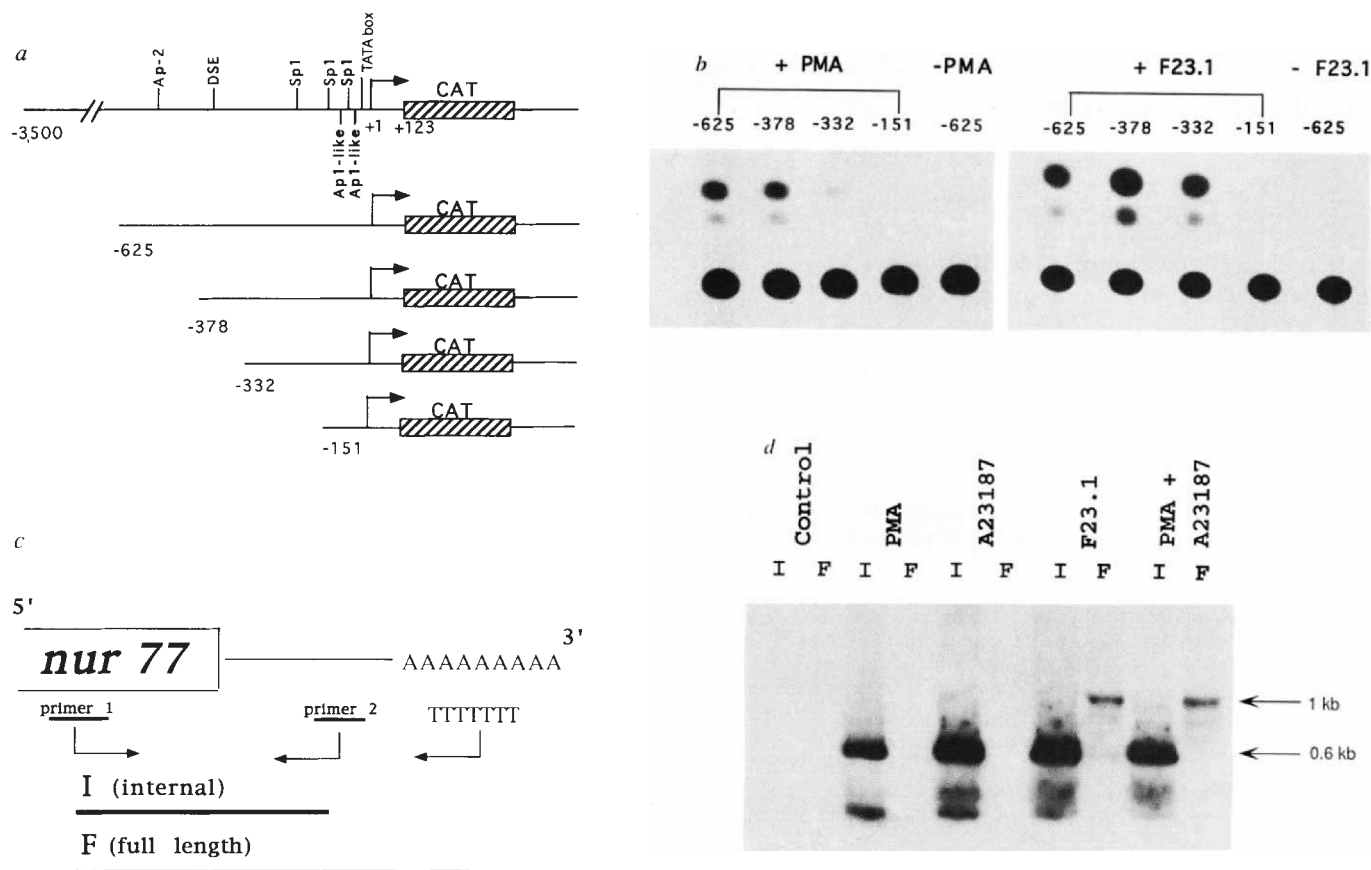


FIG. 2 Differences in transcription and mRNA structure of *nur77* following mitogenic as opposed to apoptotic signals. **a**, Diagram of deletions in 5' promoter region of *nur77* used for promoter/CAT reporter assays. Relevant control elements as described by Ryseck et al. (7) are noted. **b**, Expression of transiently transfected *nur77*/CAT constructs. DO11.10 cells were transfected with the constructs shown in **a**, treated with PMA or F23.1 and the CAT activity determined by thin-layer chromatography. **c**, The reverse transcriptase (RT)-PCR assay. **d**, DO11.10 cells were treated with PMA, the calcium ionophore A23187, anti-TCR antibody F23.1, or a combination of PMA and A23187. RNA was isolated, cDNA synthesized and analysed by RT-PCR, the products of which were separated by gel electrophoresis, Southern blotted and hybridized to a *nur77* probe. I and F correspond to internal and full-length fragments as shown in **c**. Because of sequence homology within the 3' untranslated region, the primer used to verify the presence of *nur77* generated additional products from false priming. All treatments resulted in accumulation of *nur77* mRNA, but only those initiating apoptosis accumulated significant amounts of full-length polyadenylated *nur77* mRNA.

METHODS. For construction of 5' promoter deletions, the CAT gene was ligated with the SV40 polyadenylation site into pBluescript (pBSCAT; Stratagene). A 3,500-bp *EcoRI/KpnI* fragment containing the *nur77* promoter was isolated from a *nur77* genomic clone and ligated into pBSCAT. Using an *Apal* site at -625, the -625 clone was constructed. Further nested deletions were made using this clone and *Exo I* nuclease

for digestion and *S1* nuclease to create blunt ends. All constructs were sequenced to confirm the deletion end point. Transient transfections were done using DEAE-dextran as described²². Briefly, 1×10^7 cells were transfected with 10 μg plasmid DNA in 0.5 mg ml^{-1} DEAE-dextran (M_r 500,000) at room temperature for 30 min, then cells were washed and incubated in 10 ml fresh DME/RPMI 1640 medium with 10% horse serum overnight at 37 °C. Transfectants were fed with 10 ml fresh DME/RPMI and 5 h later treated with either 10 nM PMA or placed in wells coated with F23.1 as for Fig. 1. Cells were collected 18 h after treatment and CAT activity was determined²¹. For the RT-PCR assay in **d**, 5×10^6 DO11.10 cells were treated with 10 nM PMA, or 500 nM A23187 or F23.1 (as in Fig. 1b) or with a combination of PMA and A23187. Cells were cultured as for Fig. 1 and collected 1 h after treatment. Control cells were untreated. RNA was isolated as for Fig. 1. RNA (1 μg) was converted to cDNA using a reverse transcriptase PCR kit (Perkin Elmer/Cetus) according to the manufacturer's instructions. For PCR, one primer within the coding sequence was used (5'-TGCCGAGGACCAAGAC-3') either with a primer from the 3' untranslated region (5'-CCGGTCGACGGGGTCCGTGGGC-3') or with oligo (dT) (ref. 16). Samples were denatured at 95 °C for 1 min, annealed at 42 °C for 1 min and elongated at 72 °C for 2 min for 35 cycles. Each PCR product (10 μl) was fractionated in a 1% agarose gel and transferred to Zeta-Probe membrane (BioRad) and hybridized with a ^{32}P -labelled *nur77* cDNA clone.

response to PMA requires sequences between -378 and -332, whereas responsiveness to TCR ligation maps between -332 and -151. These data indicate that signals that induce mitogenesis may act on regions of the promoter distinct from those that induce apoptosis.

In addition to the distinct transcriptional regulation of *nur77* by activation as opposed to death, there are also differences in mRNA structure. The 3' untranslated region (UTR) of *nur77* contains three repeats of the sequence AUUA⁷, a motif that may facilitate selective mRNA degradation^{14,15} or destabilization by removal of the poly(A)⁺ tail¹⁶. To examine mRNA structure, an assay based on the polymerase chain reaction (PCR) was used to verify the presence of *nur77* mRNA after various treatments, as well as to determine whether the poly(A)⁺ tail was present on the majority of transcripts. Whereas all of the treatments used in this assay could induce *nur77* (data not shown), the presence of a poly(A)⁺ tail seemed to be restricted to those conditions that resulted in cell death (Fig. 2d). These data show that the expression of an intact full-length *nur77* gene product is correlated with the commitment of a cell to undergo apoptosis. In contrast, mitogenic stimulation of cells induced *nur77* mRNA lacking a detectable poly(A)⁺ tail. The half-life of *nur77* mRNA is essentially the same after both treatments (Z.-G.L. and L.M.S., unpublished results), suggesting that the lack of a poly(A)⁺ tail may promote other functions, such as translation efficiency¹⁶. We have also observed polyadenylated *nur77* in immature thymocytes committed to die, whereas activated peripheral T lymphocytes express *nur77* lacking a poly(A)⁺ tail (data not shown).

To test the requirement of *nur77* in the induction of apoptosis in DO11.10 cells, we used a transient transfection assay. A plas-

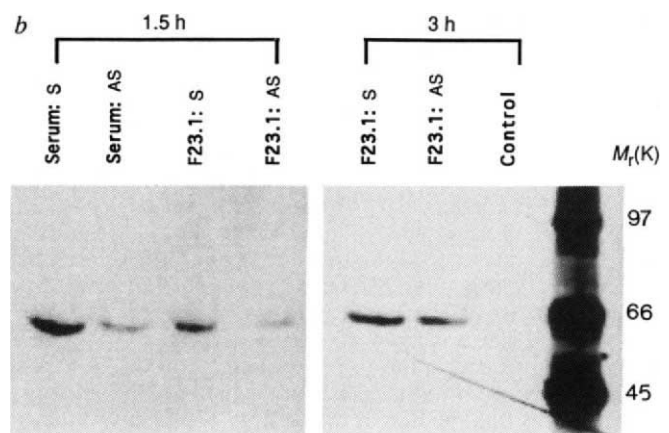
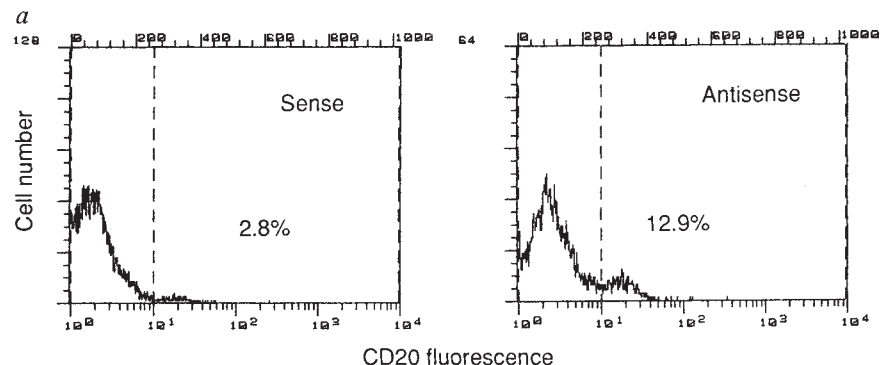
mid containing the *nur77* gene in an antisense orientation and driven by its own promoter was introduced into DO11.10 together with an expression plasmid for the cell-surface marker CD20 (ref. 17). As a control, a truncated portion of *nur77* (nucleotides 70-455) without the DNA-binding domain was co-transfected in the sense orientation with the CD20 reporter construct (Fig. 3a). Twenty-four hours after transfection, cells were stimulated with F23.1. Sixteen hours later, cells were collected and treated with anti-CD20 and ethidium bromide and analysed on a FACScan. Live cells were gated by forward scatter and ethidium bromide and CD20 profiles of live cells were analysed (Fig. 3a). After transfection with the CD20 expression plasmid alone, we observed <2% CD20⁺ cells (data not shown). When the CD20 plasmid was co-transfected with truncated *nur77* sense plasmid, less than 3% of cells were CD20⁺. But when the CD20 construct was transfected with antisense *nur77*, 12.8% CD20⁺ cells were obtained, indicating a selective survival of transfected cells expressing *nur77* antisense. These results are direct evidence that transfection of antisense *nur77* protects cells from apoptosis induced when signalled to die by TCR engagement.

To verify that *nur77* is required for the induction of apoptosis under the conditions described, we repeated these transfections, obtained CD20⁺ cells by panning, induced apoptosis with F23.1 exposure and used western blot analysis to examine expression of *nur77* protein. As shown in Fig. 3b, the presence of antisense *nur77* results in a 73% decrease in *nur77* protein within 1.5 h of induction of cell death with F23.1. As expected from our promoter mapping experiments, we found that *nur77* protein is induced by serum stimulation of DO11.10 cells. Again, antisense expression of *nur77* inhibits this induction of protein by 72%. Taken together, these data support the hypothesis that *nur77* acts as a

FIG. 3 *nur77* is required for cell death in DO11.10. DO11.10 cells were transfected with a CD20 expression plasmid (pCMVCD20) (ref. 17); cells also received a plasmid containing the *nur77* promoter driving either *nur77* in the antisense orientation or a truncated sense strand. a, 24 h after transfection, cells were incubated with F23.1 and the per cent CD20⁺ cells determined. b, Western blot of *nur77* protein following transfection with antisense (AS) or sense (S) *nur77* as in a.

METHODS. a, The truncated sense and antisense construct were derived by cloning the sense portion of *nur77* cDNA encompassing the transcriptional start site, an *EcoRI*-*SmaI* fragment containing bp 70-455 as described⁴.

This fragment was cloned into -378 CAT. The anti-sense construct was derived from the same *EcoRI*-*SmaI* fragment from the *nur77* cDNA and cloned into -378 CAT in the opposite orientation. Transfections were done as described for Fig. 2b, except that 10 µg sense or antisense plasmid was cotransfected with 5 µg pCMVCD20. 24 h after transfection, cells were passed over Percoll to remove dead cells and treated with F23.1 as described in Fig. 1 legend. Cells were collected 18 h after induction of apoptosis, washed with PBS + 2% horse serum and reacted with anti-CD20 (Pharmingen, La Jolla) + 50 ng ml⁻¹ ethidium bromide. Cells were analysed on a FACScan using FACScan Research software (Becton Dickinson). Live cells were gated by forward scatter versus FL3 (ethidium bromide) analysis and CD20 profiles of live cells determined. 30,000 cells were analysed for each condition. b, Transfections and cell preparations were done as in a, except that CD20⁺ cells were isolated by panning with 20 µg ml⁻¹ anti-CD20 before induction of apoptosis. After induction of cells with either F23.1 or serum (10% horse serum), equal numbers of cells were collected at 1.5 or 3 h, lysed with 2x Laemmli SDS buffer, boiled for 5 min and run on an 8% SDS-PAGE gel, blotted onto nitrocellulose and treated with anti-*nur77*. Loading of protein was the same in each lane, as shown by an α -tubulin antibody which detected no differences between lanes (data not shown). Western blots were developed using an Amersham ECL kit according to the manufacturer's instructions. *nur77* protein was quantified by densitometry of western blots.



critical link between TCR engagement and apoptosis in DO11.10 cells.

Our results indicate that *nur77* joins other immediate-early genes as essential mediators of apoptosis. For example, overexpression of *c-myc* in growth-arrested T cells¹⁸ or fibroblasts¹⁹ is required for induction of apoptosis. Furthermore, our data provide evidence that transcription of *nur77* mRNA is differentially regulated by mitogenic signals compared with those inducing death. They also suggest that mRNA structure may change in these two situations. The finding that genes known to be essential for mitosis are also critical in apoptosis suggests that these genes might function as branch points between the induction of proliferation and death. □

Received 27 October; accepted 3 December 1993.

1. MacDonald, H. R. & Lees, R. K. *Nature* **343**, 642–644 (1990).
2. Swat, W., Ignatowicz, L., von Boehmer, H. & Kiselow, P. *Nature* **351**, 150–153 (1991).
3. Smith, C. A., Williams, G. T., Kingston, R., Jenkinson, E. J. & Owen, J. J. T. *Nature* **337**, 181–184 (1989).
4. Hazel, T. G., Nathans, D. & Lau, L. F. *Proc. natn. Acad. Sci. U.S.A.* **85**, 8444–8448 (1988).
5. Milbrandt, J. *Neuron* **1**, 183–188 (1988).
6. Varnum, B. C., Lim, R. W. & Herschman, H. R. *Oncogene* **4**, 1263–1265 (1989).

7. Ryseck, R.-P., MacDonald-Bravo, H., Mattei, M.-G., Ruppert, S. & Bravo, R. *EMBO J.* **8**, 3327–3335 (1989).
8. Ucker, D. S., Ashwell, J. D. & Nickas, G. J. *Immun.* **143**, 3461–3469 (1989).
9. Murphy, K. M., Heimberger, A. B. & Loh, D. Y. *Science* **250**, 1720–1722 (1990).
10. Schwartz, L. M., Kosz, L. & Kay, B. K. *Proc. natn. Acad. Sci. U.S.A.* **87**, 6594–6598 (1990).
11. Haskins, K., Kubo, R., White, J., Pigeon, M., Kappler, J. & Marrack, P. J. *exp. Med.* **157**, 1149–1169 (1983).
12. Ashwell, J. D., Cunningham, R. E., Noguchi, P. D. & Hernandez, D. J. *exp. Med.* **165**, 173–194 (1987).
13. Wyllie, A. H. *Nature* **284**, 555–556 (1980).
14. Shaw, G. & Kamen, R. *Cell* **46**, 659–667 (1986).
15. Wilson, T. & Treisman, R. *Nature* **336**, 396–399 (1988).
16. Monroe, D. & Jacobson, A. *Molec. cell. Biol.* **10**, 3441–3455 (1990).
17. Zhu, L. et al. *Genes Dev.* **7**, 1111–1125 (1993).
18. Shi, Y. et al. *Science* **257**, 212–214 (1992).
19. Evan, G. I. et al. *Cell* **69**, 119–128 (1992).
20. Shi, Y. et al. *J. Immun.* **144**, 3326–3333 (1990).
21. Leo, O., Foo, M., Sachs, D. H., Samelson, L. E. & Bluestone, J. A. *Proc. natn. Acad. Sci. U.S.A.* **84**, 1374–1378 (1987).
22. Auebel, F. M. et al. *Current Protocols in Molecular Biology* (Wiley Interscience, New York, 1987).

ACKNOWLEDGEMENTS. Z.-G.L. and S.W.S. contributed equally to the work presented in this letter. We thank K. Murphy and D. Loh for mRNA to construct and screen the library; R. Horvitz for support, advice and discussion; E. Harlow and S. van den Heuvel for the CD20 plasmid; J. Milbrandt and L. Lau for *α-nur77*; A. Ponce de Leon for help with Fig. 3b; and R. Goldsby, D. Green, K. Murphy, L. Berg, G. Baughman, S. J. Smith and members of our laboratories for discussion. We also acknowledge the original observation of the induction of *nur77* in T-cell hybrids by B. Stanger. This work was supported by grants from the NIH to B.A.O. and L.M.S.

Prediction of bone density from vitamin D receptor alleles

Nigel A. Morrison, Jian Cheng Qi, Akifumi Tokita, Paul J. Kelly, Linda Crofts, Tuan V. Nguyen, Phillip N. Sambrook & John A. Eisman

Bone and Mineral Research Division, Garvan Institute of Medical Research, St Vincent's Hospital, Sydney, New South Wales 2010, Australia

BONE density achieved in early adulthood is the major determinant of risk of osteoporotic fracture. Up to 60% of women^{1,2} suffer osteoporotic fractures as a result of low bone density², which is under strong genetic control^{3–6} acting through effects on bone turnover^{7,8}. Here we show that common allelic variants in the gene encoding the vitamin D receptor⁹ can be used to predict differences in bone density, accounting for up to 75% of the total genetic effect on bone density in healthy individuals. The genotype associated with lower bone density was overrepresented in postmenopausal women with bone densities more than 2 standard deviations below values in young normal women. The molecular mechanisms by which bone density is regulated by the vitamin D receptor gene are not certain, although allelic differences in the 3' untranslated region may alter messenger RNA levels. These findings could open new avenues to the development and targeting of prophylactic interventions. It follows that other pathophysiological processes considered to be subject to complex multifactorial genetic regulation may also be modulated by a single gene with pleiotropic transcriptional actions.

We have shown that common allelic variation in the vitamin D receptor locus (*VDR*) can be used to predict bone turnover¹⁰. We therefore examined the effect of *VDR* alleles on bone density using twins as a model in which within-pair comparisons eliminate age and various cohort confounders. The study comprised 250 normal healthy caucasian twins from Sydney and Melbourne, 70 monozygotic (MZ) and 55 dizygotic (DZ) twin pairs, with mean age 45 ± 13 (MZ) and 44 ± 11 (DZ) yr, range 17–70 yr. Seven MZ and six DZ pairs were male and the female twins were premenopausal or concordant for years since menopause. Bone density was measured at the lumbar spine and proximal femur by dual-photon or dual energy X-ray absorptiometry (Lunar Corporation, Madison)^{11,12}.

Monozygotic twins share 100% and DZ twins share 50% of their genes. Thus if a gene contributes to an observed genetic effect, dizygotic twins concordant for alleles of that gene should be more similar to each other than discordant twins and comparable to monozygotic rather than to other dizygotic twins. In both the clinically important regions of lumbar spine and proximal femur, the within-pair difference in bone density was significantly less in concordant than in discordant DZ twins (Fig. 1). Importantly, the within-pair difference in bone density for lumbar spine in concordant DZ twins was not significantly different from that in MZ twins, and in both was less than that in discordant DZ twins ($P < 0.0001$). Similar but weaker effects in the proximal femur are consistent with stronger environmental influences in this region^{12–14}. *VDR* genotype was the strongest predictor at the lumbar spine ($P = 0.0002$) and the trochanteric region ($P = 0.02$) after controlling for potential confounding by height and weight.

In view of the co-dominant effect of *VDR* alleles on bone turnover¹⁰, a co-dominant effect on bone density was expected. Using modified sib-pair linkage analysis^{15,16}, in which a significant correlation between the squared difference in a trait and the proportion of genes identical-by-state within a sibling pair indicates genetic linkage, the *VDR* alleles were co-dominant with 1.5–2.5-fold greater within-pair differences for discordant twins (Fig. 1). In 21 of 22 dizygotic twin pairs discordant for the *VDR* alleles, the *b* allele was associated with higher bone density (Fig. 2). When all twins were analysed by ANOVA (mixed model with random error), the *bb* genotype was associated with higher bone density with a clear co-dominant effect (Fig. 2). When the analysis was limited to females, all bone sites showed significant effects of *VDR* genotype; lumbar spine, $P = 0.0000036$; femoral neck, $P = 0.06$; Ward's triangle, $P = 0.05$; trochanteric region, $P = 0.019$.

To confirm the effect in another population, a further 311 unrelated healthy women, aged 52.5 ± 13.5 yr from the Sydney region were examined, including 207 postmenopausal women of 0.5 to 46 (mean 11.3) years post-menopause. Genotype was an independent predictor at both lumbar spine and femoral neck in multiple regression including height (cm), weight (kg), age, years post-menopause (ypm): as lumbar spine

$$\text{BMD} = 0.19 - 0.004 \text{ age} + 0.043 \text{ genotype} + 0.005 \text{ cm} \\ + 0.003 \text{ kg} - 0.004 \text{ ypm}$$

($R^2 = 0.34$, with P values of 0.0009, 0.0012, 0.001, 0.011 and 0.01

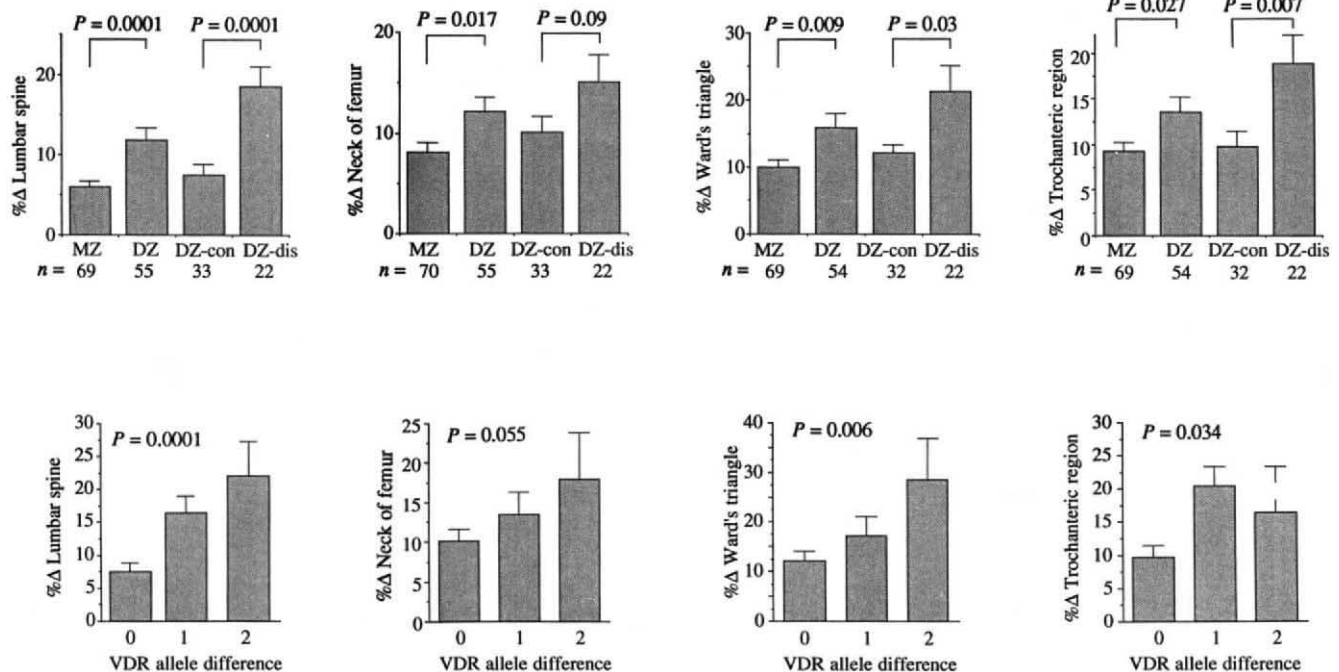


FIG. 1 Bone density differences in lumbar spine and proximal femur of twin pairs with respect to zygosity and concordance for VDR alleles. Dizygotic twins concordant for the VDR alleles are not different from the MZ twins at any site; discordant DZ twins are significantly different from both of these groups, with 48–90% of the genetic effect explained by VDR alleles (upper panels). Regression of allele-dose effect was significant at lumbar spine ($P=0.0001$), Ward's triangle ($P=0.006$) and trochanteric region ($P=0.03$), and borderline at the femoral neck ($P=0.055$) (lower panels). Using the sib-pair variance approach^{15,16}, significant relationships were observed between the squared difference in bone density within each twin pair and concordance for the VDR alleles at the lumbar spine ($r=0.42$, $P=0.001$), femoral neck ($r=0.28$, $P=$

0.04), Ward's triangle ($r=0.33$, $P=0.01$) and trochanteric regions ($r=0.26$, $P=0.05$). Another developmental transcriptional activator, the retinoic acid receptor- α (on chromosome 17q21; ref. 21), was not predictive at any site. The polymorphic sites for *BsmI* and *Apal* (ref. 10) are in the region of the gene from exon 7 to the 3' UTR. No polymorphic sites were in the coding region except for a T-for-C synonymous coding change in exon 9, changing ATT to ATC (isoleucine), creating a *TaqI* polymorphism. PCR of the DNA sequence flanking the *BsmI* site was used to facilitate genotyping of subjects with a Corbett FTS-1 Thermal Sequencer (Corbett Research, NSW, Australia) with primers 5'-CAACC-AAGACTACAAGTACCGTCAGTGA-3' and 5'-AACCAGCGGAAG-AGGTCAAGGG-3'.

respectively) and femoral neck

$$\text{BMD} = 0.26 - 0.005 \text{ age} + 0.021 \text{ genotype} + 0.003 \text{ cm} \\ + 0.004 \text{ kg} - 0.003 \text{ ypm}$$

($R^2=0.50$, with P values of 0.0001, 0.025, 0.006, 0.0001 and 0.013 respectively).

Bone density at both lumbar spine and femoral neck was analysed between genotypes for the so-called 'fracture threshold', two standard deviations (s.d.) below the mean of young

normal values¹. Bone density at the lumbar spine in the *BB* genotype group would reach this level 18.4 yr after the menopause compared with 29 yr for the *bb* genotype, with heterozygotes intermediate at 22 yr (Fig. 3). Using age the intercepts were: *BB*, 65 yr; *Bb*, 69 yr; *bb*, 76 yr (Fig. 3). As femoral neck bone density declines linearly with age¹, femoral neck bone density was analysed between genotypes according to age. In the *BB* genotype, the mean value reached the -2 s.d. level at age 63 yr compared with 71 yr in the *bb* genotype group (data not shown).

A genetic test for predisposition to low bone density could

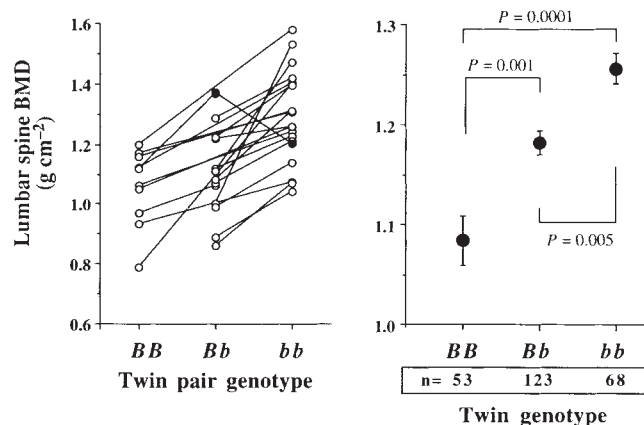


FIG. 2 Higher bone density is associated with the *b* allele of the VDR gene. Lumbar spine bone densities of dizygotic twins discordant for VDR alleles ($n=22$) are plotted as twin and co-twin according to genotype, with lines joining each twin pair (left panel). In 21 of 22 associated twin pairs, the twin carrying one or two extra copies of the *b* allele has the higher bone density. A single pair of young male twins exhibited the reverse situation. Bone mineral density (mean $\pm$ s.e.m.) is shown for all twins according to genotype regardless of zygosity (right panel). It is clear that the *BB* genotype is associated with lower mean bone density. The magnitude of the effect can be seen in relation to the standard deviation of an age-matched population of $\sim 0.11 \text{ g cm}^{-2}$. The significance of difference between groups was calculated by the mixed-model ANOVA, incorporating a fixed effect for genotype and random effects for other twin pair-specific factors. At the lumbar spine the effect of genotype was significant at $P=0.000054$, and at the femoral neck $P=0.038$. The P values displayed are for pairwise Student's *t*-tests between the genotypes. A similar significant effect was observed for the Ward's triangle ($P=0.033$) but not the trochanteric region ($P=0.1$) by mixed model random effect ANOVA. When the analysis was limited to females, P values were: lumbar spine, $P=0.0000036$; femoral neck, $P=0.06$; Ward's triangle, $P=0.05$; trochanteric region, $P=0.019$.

be invaluable for prediction of risk at times when preventative approaches are being considered, for instance in adolescence. Although allele typing cannot account for all factors, in those postmenopausal women with bone density more than 2 s.d. below the mean of young normal women, there was overrepresentation of the *BB* genotype ($\chi^2 P < 0.001$). Of females over 60 yr, 74% of the *BB* genotype were below the fracture threshold, and 61% of the *bb* genotype were above it. The magnitude of the *VDR* gene effect could translate to a 10-yr difference in age of incidence of spine fractures between the alternate homozygote groups (Fig. 3), comparable to the strongest known factor, menopausal oestrogen deficiency. Similarly, as the incidence of hip fractures doubles every 4–5 years², the age difference of 8 yr for femoral neck bone density could translate to a fourfold risk of hip fracture. These differences for both hip and spine fractures are potentially of major clinical importance.

The alleles are in strong linkage disequilibrium with several other polymorphisms, including a synonymous codon change in exon 9 which has 97% concordance of genotype with the *BsmI* restriction-fragment length polymorphism (RFLP), as well as

other sequence changes in the 3' untranslated region (UTR). In minigene analysis of the 3' UTR from the two most frequent haplotypes, BAt and baT, in a kidney and bone cell line (Fig. 4) there was substantially greater luciferase reporter gene activity from the BAt haplotype (1.44:1 and 1.38:1), consistent with enhanced gene transcription or mRNA stability. Preliminary analysis using reverse transcriptase-polymerase chain reaction (RT-PCR) in heterozygotic cell lines gave similar differences in receptor mRNA levels (data not shown). The biochemical basis of the *VDR* gene effect on bone density is uncertain, however, and needs further investigation. The active hormonal form of vitamin D (calcitriol) is a central regulator of bone and calcium homeostasis, modulating intestinal calcium absorption, bone formation, recruitment and function of bone-resorbing cells (osteoclasts), parathyroid hormone production and renal vitamin D activation. Small differences in expression or function of steroid hormone receptors, which amplify hormonal signals through networks of genes, can mediate large differences in end-product target genes. The *VDR*, acting through a vitamin D response element^{17–19}, is therefore a good candidate for a prime

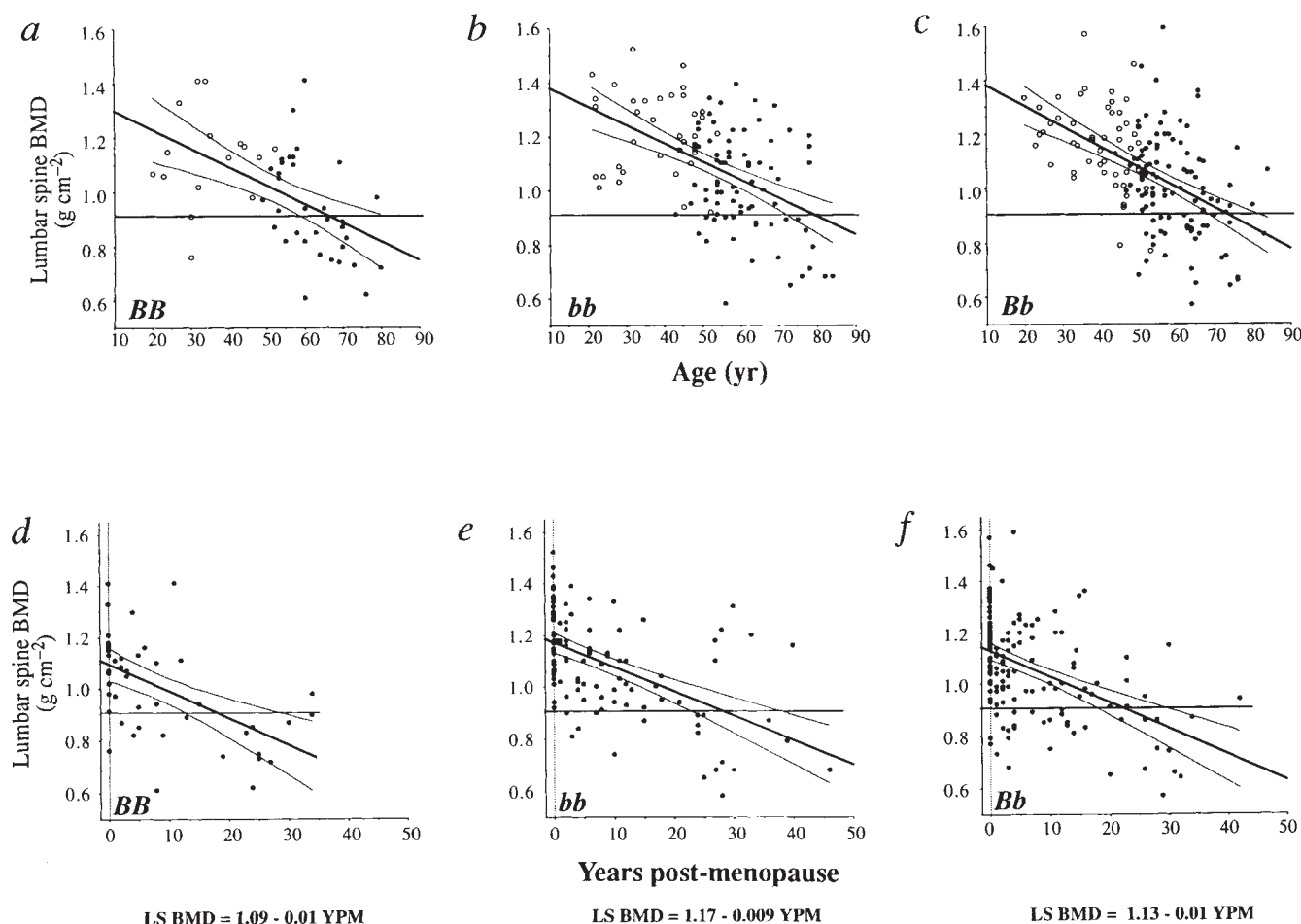
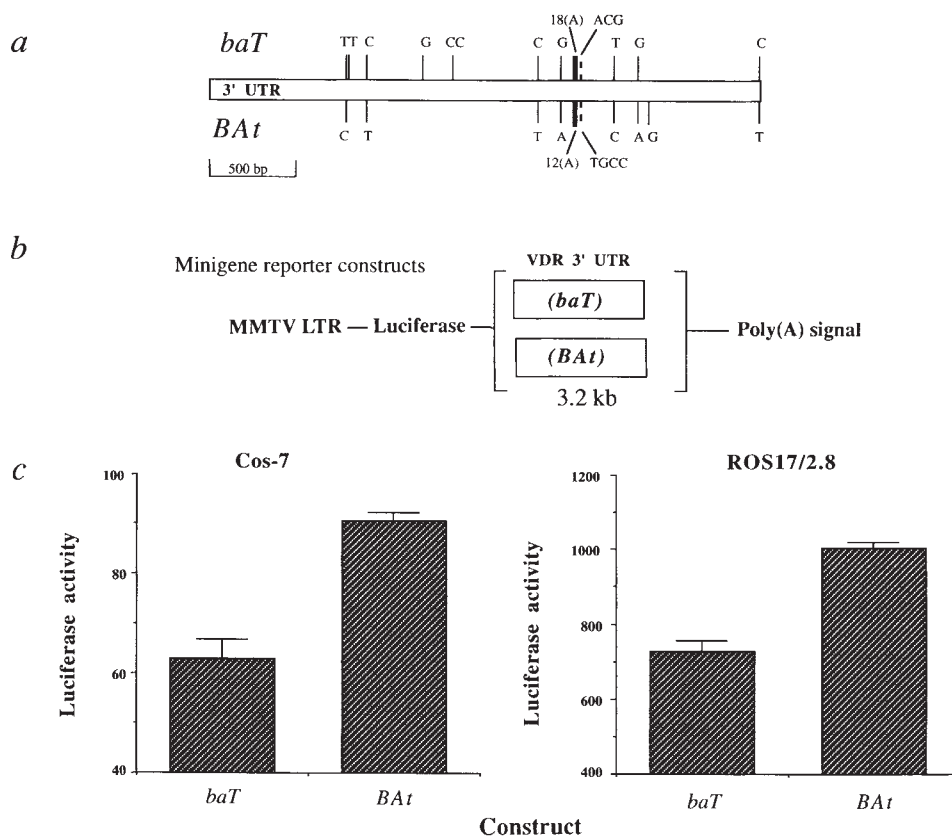


FIG. 3 Vitamin D receptor genotype and lumbar spine bone density in unrelated normal women with respect to age and years post-menopause. The bone density in the lumbar spine is plotted according to *VDR* genotype and age-related (a, b and c) and years since menopause-related change (d, e, f). The regression lines and 95% confidence limits of the mean of bone density for the three genotypes (*BB*, *Bb*, *bb* as marked) are shown. For the lumbar spine, genotypes *BB* and *bb* had significantly different intercepts of the regression with a bone density 2 s.d. below the mean of young normal female values at $P < 0.04$. The mean of young normal females was derived from a Sydney study population, aged between 35 and 40 yr, and was 1.20 ± 0.144 s.d. g cm^{-2}

for lumbar spine BMD, $n = 70$. The -2 s.d. value was 0.91 g cm^{-2} and is shown as a horizontal line in all graphs. Premenopausal (open circles) and postmenopausal women (filled circles) are shown. Regression with years post-menopause (coding premenopausal women as zero years post-menopause) gave similar results with intercepts of 18.4, 22 and 29 years post-menopause for the *BB*, *Bb* and *bb* genotypes respectively. Similar analysis of the age-related changes in femoral neck bone density gave intercept values (for a value of 0.7 g cm^{-2} as 2 s.d. below young normal) of 16, 21 and 25 years post-menopause, but these differences were not significant.

FIG. 4 Minigene analysis of potential effect of common alleles of the 3' UTR of the vitamin D receptor gene on transcriptional activity and mRNA stability. This showed that allelic sequence variation in the 3' UTR of the vitamin D receptor, assessed in a heterologous gene system, can alter either or both of the transcriptional efficiency of the gene or the stability of the transcribed mRNA. **a**, The 3' UTR from two individuals homozygous for *BsmI*, *Apal* and *TaqI* RFLPs, *BBAAtt* and *bbaaTT* genotypes, which define the two most frequent haplotypes, *BAAt* and *baT*, of the vitamin D receptor gene¹⁰, were amplified using *Tth* polymerase (Toyobo Inc). Between the two haplotypes, shown above and below the 3' UTR, there were a number of sequence differences, including point sequence changes, insertion/deletions, one length polymorphism and a 3/4-bp substitution involving 13 distinct sites. By contrast, in the adjacent 2.4 kb of coding and intronic sequence, there was only one synonymous codon change in exon 9 (T to C) and four intronic point differences, including those for the *BsmI* and *Apal* polymorphisms. **b**, The 3.2-kb 3' UTRs from exon 9 of the two common haplotypes were inserted into otherwise identical expression vectors, downstream of the firefly luciferase gene (Promega) in a minigene construct driven by the mouse mammary tumour virus long terminal repeat (MMTV LTR; Pharmacia). **c**, The ratio of luciferase activity of the minigenes transfected into COS-7 (green monkey kidney) cells, after correction for transfection efficiency



with pTKCAT, was 1.44 ± 0.02 ($P = 0.0000012$, $n = 16$). In rat osteosarcoma cells (ROS 17/2.8) a similar ratio of 1.38:1 ($P < 0.001$, $n = 11$) was found. In both cases, the *BAAt* haplotype 3' UTR supported greater luciferase activity.

regulator of bone and calcium homeostasis and bone density. In support of a physiological mechanism involving the VDR, serum calcitriol levels were higher in the *BB* genotype: 134 ± 42 pM for *BB*, $n = 29$; 104 ± 30 pM for *Bb*, $n = 59$; 99 ± 40 pM for *bb*, $n = 29$; $P = 0.0008$ by mixed model random error ANOVA. Functionally distinct *VDR* alleles might contribute to the differences in bone and calcium homeostasis and bone mass between different ethnic groups such as American blacks, whites and Hispanics²⁰.

Irrespective of the molecular biological or physiological mechanisms, our data have identified a gene locus associated with a major part of the strong genetic effect on bone density and indeed more than half of the adjusted population variation in bone density. Use of this genetic marker should allow earlier intervention in those at increased risk of osteoporosis, provide insight into the physiological mechanisms of the wide population variance in bone density, and open the way to development of specific targeted therapy. □

Received 17 September; accepted 1 December 1993.

- Riggs, B. L. & Melton, L. J. *New Engl. J. Med.* **314**, 1676–1686 (1986).
- Nguyen, T. V. et al. *Br. Med. J.* **307**, 1111–1115 (1993).
- Smith, D. M., Nance, W. E., Kang, K. W., Christian, J. C. & Johnston, C. C. *J. clin. Invest.* **52**, 2800–2808 (1973).
- Moller, M. et al. *Calcif. Tissue Res.* **25**, 197–201 (1978).
- Pocock, N. A. et al. *J. clin. Invest.* **80**, 706–710 (1987).
- Dequeker, J. et al. *Bone* **8**, 207–209 (1987).
- Seeman, E. et al. *New Engl. J. Med.* **320**, 554–558 (1989).
- Kelly, P. J., Eisman, J. A. & Sambrook, P. N. *Osteoporosis Int.* **1**, 1–5 (1990).
- Kelly, P. J. et al. *J. clin. Endocrin. Metab.* **72**, 808–814 (1991).
- Morrison, N. A., Yeoman, R., Kelly, P. J. & Eisman, J. A. *Proc. natn. Acad. Sci. U.S.A.* **89**, 6665–6669 (1992).
- Pocock, N. A. et al. *Med. J. Aust.* **146**, 293–297 (1987).
- Pocock, N. A. et al. *J. Bone Min. Res.* **3**, 601–604 (1988).
- Pocock, N. A., Eisman, J. A., Yeates, M. G., Sambrook, P. N. & Eberl, S. J. *J. clin. Invest.* **78**, 618–621 (1986).

- Pocock, N. A. et al. *J. Bone Min. Res.* **4**, 441–448 (1989).
- Haseman, J. K. & Elston, R. C. *Behav. Genet.* **1**, 11–19 (1970).
- Haseman, J. K. & Elston, R. C. *Behav. Genet.* **2**, 3–19 (1972).
- Morrison, N. A. et al. *Science* **246**, 1158–1161 (1989).
- Demay, M. B., Roth, D. A. & Kronenberg, H. M. *J. biol. Chem.* **264**, 2279–2282 (1989).
- Kerner, S. A., Scott, R. A. & Pike, J. W. *Proc. natn. Acad. Sci. U.S.A.* **86**, 4455–4459 (1989).
- Bell, N. H. et al. *J. clin. Invest.* **76**, 470–473 (1985).
- Arveiler, B., Petkovich, M., Mandel, J. L. & Chambon, P. *Nucleic Acids Res.* **16**, 6252 (1988).

ACKNOWLEDGEMENTS. The majority of the twins were recruited through the Australian NH and MRC twin registry. We thank E. Seeman for providing additional twins, J. Birmingham and D. Fontana for assistance, and P. Chambon for the cDNA probe for retinoic acid receptor studies. This work was supported by the NH and MRC of Australia and the NIH (USA). J.C.C. was supported by the K. C. Wang Foundation (Hong Kong) and the Garvan Research Foundation (Sydney).

A new chromosomal protein essential for mitotic spindle assembly

Jing-Ping Yeo, Frank Alderuccio & Ban-Hock Toh*

Department of Pathology and Immunology,
Monash University Medical School, Melbourne,
Victoria 3181, Australia

ASSEMBLY of the mitotic spindle, the machinery responsible for chromosomal segregation, is regulated by Cdc2 kinase¹⁻⁵ and requires mitotic chromatin⁵⁻⁸. However, the molecular identity of the kinase substrate and chromatin factor is unknown. Here we have cloned a human complementary DNA encoding an evolutionarily conserved chromosomal protein of relative molecular mass 47,000 (M_r 47K) which has three consensus motifs for Cdc2 kinase-mediated phosphorylation⁹. The protein is phosphorylated only during mitosis and is associated with polypeptides having M_r s of 31K, 67K and 200K. Mitotic arrest is induced by antisense messenger RNA or by affinity-purified autoantibody. In the arrested cells, the chromosomes remain unsegregated and the mitotic spindle is absent. We propose that the chromosomal protein is activated by phosphorylation at the interphase/mitosis transition by Cdc2 kinase, and that the protein, alone or as a complex, is a previously unidentified Cdc2 kinase substrate and chromatin factor necessary for spindle assembly.

The autoimmune serum from a patient with discoid lupus erythematosus reacted with chromosomes of mitotic Hep-2 cells, but not with nuclei of interphase cells (Fig. 1). Similar reactions were seen in monkey, mouse and marsupial cells (data not shown). The serum immunoblotted a 47K component in human, monkey, mouse and chicken cells (Fig. 2a), indicating that the molecule is evolutionarily conserved. The 47K component, designated RMSA-1 (regulator of mitotic spindle assembly), was present in the nuclear (Fig. 2b, lane 2) but not the cytoplasmic (Fig. 2b, lane 3) fraction, and has a neutral isoelectric point (pI) in interphase cells (I, Fig. 2c). Further, the 47K protein was found only in HeLa cell karyoplasts free of centrosomes, the cytoskeleton and other cytoplasmic components (Fig. 2b, lanes 4, 6), confirming that it is a nuclear protein. Although immunofluorescence was observed only during mitosis, RMSA-1 was present throughout the cell cycle (Fig. 2d), suggesting differential immunofluorescence reactivity of antisera with phosphorylated and non-phosphorylated protein in fixed preparations.

RMSA-1 protein was phosphorylated in metaphase-arrested (Fig. 2e, lane 6) but not in interphase cells (Fig. 2e, lane 5), although equivalent RMSA-1 levels were present in immunoprecipitates from cells labelled with ³⁵S-methionine/cysteine (Fig. 2e, lanes 3, 4). Similarly, phosphorylated RMSA-1 was immunoprecipitated from synchronized HeLa cells in G2/M but not from cells in G0/G1 or S phases (data not shown). Further, multiple spots, consistent with phosphorylation of RMSA-1, were observed in two-dimensional immunoblots of metaphase cells (M, Fig. 2c). These results indicate that RMSA-1 is phosphorylated only during mitosis.

31K, 67K and 200K molecules were consistently co-precipitated with RMSA-1 from ³⁵S-methionine-labelled interphase and metaphase-arrested HeLa cells (Fig. 2e, lanes 3, 4). These molecules are tightly associated with RMSA-1, as they were retained despite washing the immunoprecipitates with buffers containing 0.7 M sodium chloride, conditions that extracted the complex from chromatin. The associated polypeptides were also co-phosphorylated in mitosis (Fig. 2e, lane 6), suggesting that they constitute a functional complex with RMSA-1.

We cloned a human cDNA encoding a protein of 418 amino

acids with a predicted M_r of 48.24K and a pI of 6.33, properties consistent with those of RMSA-1. The deduced amino-acid sequence (Fig. 3c) contains conserved nuclear localization signals¹⁰, consistent with our observations that RMSA-1 is a nuclear protein. The amino-acid sequence also contains three ST/P consensus motifs for phosphorylation by Cdc2 kinase⁹, although the proposed phosphorylation sites do not contain the flanking basic amino acids found in many sites phosphorylated by the kinase¹¹. Northern blot analysis revealed a single 3-kilobase (kb) mRNA species (Fig. 3b, lane 1), in agreement with the size of the cDNA. The bacterial fusion protein containing the carboxy terminus of RMSA-1 has an M_r consistent with the predicted combined M_r of the 3K β -galactosidase fragment and partial RMSA-1. Further, the fusion protein reacted with the autoimmune serum (Fig. 2f, lanes 2, 4). Autoantibody, affinity-purified by elution from the cDNA clone, reacted with chromosomes by immunofluorescence, immunoblotted a 47K molecule (Fig. 2f, lane 6) and immunoprecipitated RMSA-1 and associated molecules. No reactions were observed with serum eluted from a non-immunoreactive clone (Fig. 2f, lane 5). *In vitro* transcription and translation of the cDNA produced a 47K product (Fig. 2g, lane 2) immunoprecipitated by the autoimmune serum (Fig. 2g, lane 5). Taken together, the data provide compelling evidence that the cDNA encodes RMSA-1.

We regulated production of antisense transcripts in transfected mouse fibroblasts with a synthetic mouse metallothionein promoter^{12,13}. After 6 h treatment with 100 μ M Zn²⁺, transfected cells exhibited size variations, cytoplasmic granules and blunted processes (Fig. 4A, panel c); at 12 h, cells were rounded (Fig. 4A, panel d) and at 24 h, most rounded cells had become

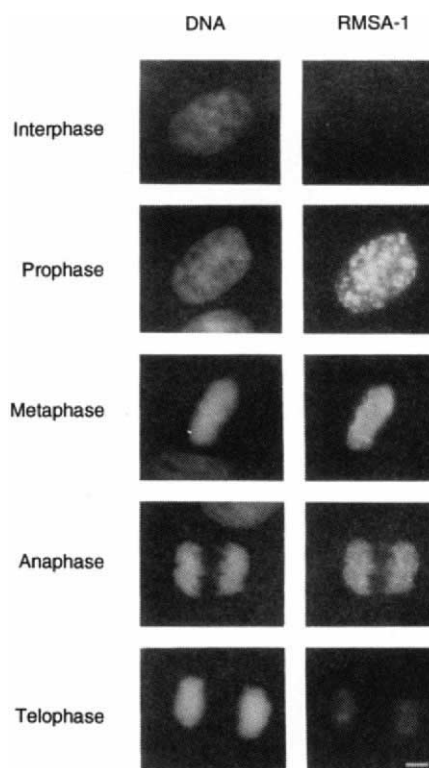


FIG. 1 Immunofluorescence localization of RMSA-1 autoantigen to chromosomes of mitotic cells. Human Hep-2 cells (Kallestad) were incubated with autoimmune serum from a patient with discoid lupus erythematosus, diluted 1:200 in phosphate-buffered saline (PBS) and traced with fluorescein isothiocyanate (FITC)-labelled anti-human immunoglobulin (Wellcome). The cells were washed and counterstained with 5 μ M Hoechst 33342 DNA dye. Note intense immunofluorescence of mitotic cells in prophase, metaphase and anaphase, with weaker immunofluorescence in telophase and absence of immunofluorescence in interphase cells. Scale bar, 2 μ m.

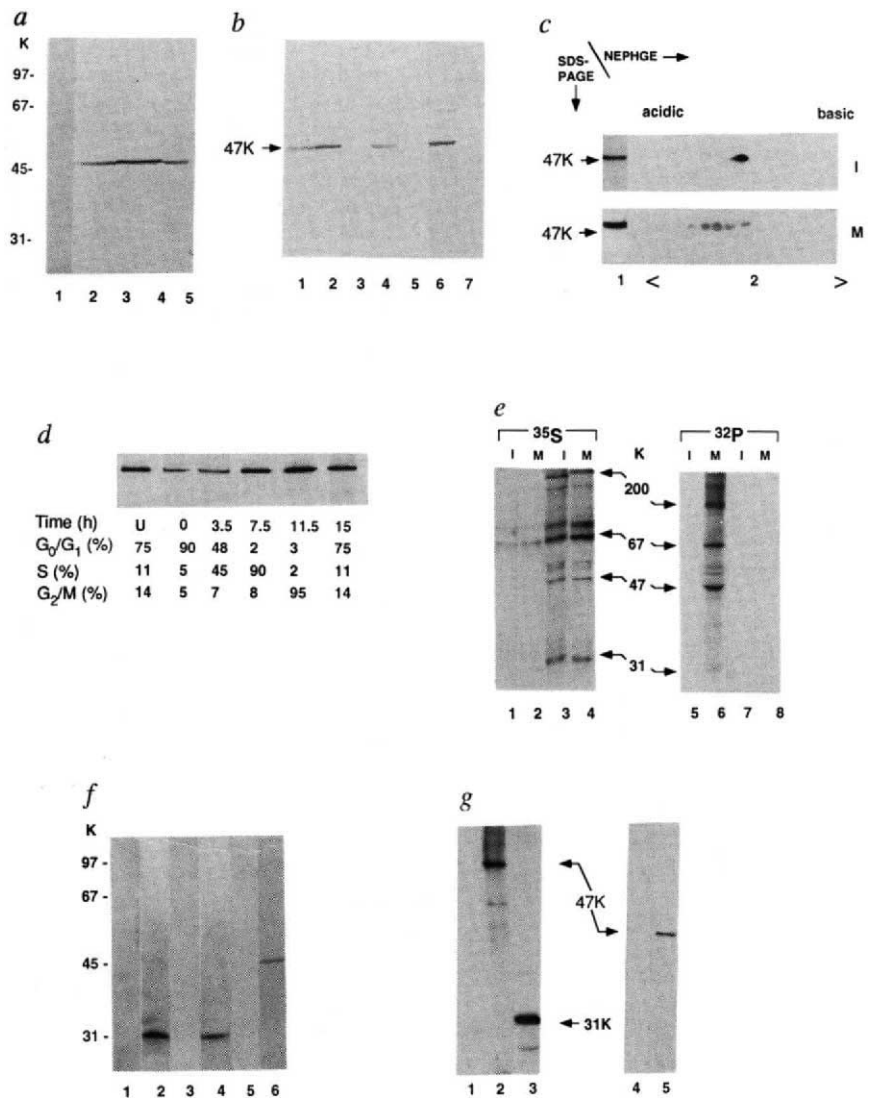
* To whom correspondence should be addressed.

detached (Fig. 4A, panel e). These morphological changes were not observed in antisense-transfected cells without zinc, zinc-treated sense-transfected cells (Fig. 4A, panel a), or untransfected cells with or without zinc. Cell viability in control cells remained unchanged with or without zinc. After 24 h zinc treatment, most antisense-transfected cells were arrested in mitosis, displaying chromosomes that had not moved to the cell equator (Fig. 4B, panel a). These mitotically arrested cells did not react with anti-RMSA-1 antibody by immunofluorescence or by

immunoblotting (Fig. 4C), indicating that antisense mRNA had inhibited RMSA-1 expression. The antisense-transfected cells treated with zinc showed progressive increase in cells blocked at G2/M with a corresponding decrease in cells at G1 and S, so that by 24 h treatment most cells were blocked at G2/M (Fig. 4D). The morphological changes are initially reversible, with cells reverting to a normal phenotype 3 h after zinc removal from the culture medium. However, after 24 h the changes are irreversible, with cell death from apoptosis (data not shown).

FIG. 2 Biochemical characterization of RMSA-1. **a**, The autoimmune serum immunoblots a 47K antigen in HeLa (lane 2), mouse 3T3 (lane 3), monkey Cos (lane 4) and chicken fibroblast cells (lane 5); normal human serum does not react with HeLa extracts (lane 1). **b**, Immunoblot reactivity with HeLa antigen (lane 1) is present in the nuclear (lane 2) and crude and purified karyoplast fractions (lanes 4 and 6, respectively), but not in the crude cytoplasmic fraction (lane 3) nor in the non-karyoplast cytoplasmic fractions (lanes 5, 7). **c**, Immunoblots of interphase HeLa cells (I, top panel) or from metaphase-arrested cells (M, bottom panel), separated by SDS-PAGE (lane 1) and two-dimensional SDS-PAGE (panel 2); note multiple 47K spots in metaphase-arrested cells consistent with RMSA-1 phosphorylation. NEPHGE, non-equilibrium pH gel electrophoresis. **d**, RMSA-1 presence throughout the cell cycle demonstrated by immunoblots with synchronized HeLa cells after release from double thymidine block. Percentages of cells at G0/G1, S and G2/M are indicated. Equivalent protein amounts (100 µg) were loaded on each lane. U, unsynchronized HeLa cells. **e**, Affinity-purified autoantibody (see legend for **f**) immunoprecipitates the 47K antigen and 31K, 67K and 200K molecules from ^{35}S -methionine/cysteine-labelled, interphase (I, lane 3) and metaphase-arrested (M, lane 4) HeLa cells; lanes 1, 2, normal human IgG immunoprecipitates. The same molecules were immunoprecipitated from ^{32}P -orthophosphate-labelled, metaphase-arrested, HeLa cells (M, lane 6), but not from interphase cells (I, lane 5); lanes 7, 8, normal human IgG immunoprecipitates. **f**, The autoimmune serum (lane 2, 4) immunoblots a 31K fusion protein generated by pJPL4 (lane 2) and pJPL41.6 (lane 4); lanes 1, 3, normal human serum controls. Autoantibody, affinity-purified from λ JPL4 fusion protein, reacts with a 47K antigen in HeLa cells (lane 6); material eluted from a non-immunoreactive clone is unreactive (lane 5). **g**, The *in vitro* transcription and translation product of pJPZ1L4 yields a major 47K product (lane 2) immunoprecipitated by autoimmune serum (lane 5), but not by normal human serum (lane 4). Controls are no RNA (lane 1) and α -factor RNA (lane 3).

METHODS. **a**, Proteins (10 µg per lane), separated by SDS-PAGE and transferred to nitrocellulose membrane, were immunoblotted with human sera²². **b**, Crude nuclear and cytoplasmic fractions were derived²³ from HeLa cells. Monolayers (1×10^7) were extracted on ice for 1 min with buffer containing 10 mM HEPES, pH 7.4, 0.3 M sucrose, 50 mM NaCl, 2.5 mM MgCl_2 , 1 mM CaCl_2 , 50 µM leupeptin and 0.5% Triton X-100; the cytosolic extract was collected. Monolayers were extracted for 15 min at 4 °C in buffer containing 0.1 M Tris-HCl, pH 8.0, 5 mM KCl, 1 mM CaCl_2 , 0.5 mM MgCl_2 , 0.5% NP40, 50 mM leupeptin and 0.1 M, 0.5 M, 0.7 M or 1 M NaCl. Cell fragments were scraped off, pelleted and the supernatant kept as the nuclear extract. Data shown are for 0.1 M NaCl extraction. Karyoplasts were prepared as described⁵. **c**, HeLa lysates (100 µg) from interphase or metaphase-arrested cells were processed for two-dimensional immunoblotting²⁴. **d**, HeLa cells were synchronized by double thymidine block²⁵, released from block and sampled for flow cytometry and immunoblotting. **e**, HeLa cells were metaphase-arrested by 20 h incubation with nocodazole (0.2 µg ml⁻¹) at 37 °C. Metaphase-arrested and interphase cells were radiolabelled



with 0.5 mCi ^{35}S -methionine/cysteine in methionine and cysteine-free medium or with 1 mCi ^{32}P -orthophosphate in phosphate-free culture medium for 8 h at 37 °C. Labelled cells were lysed and nuclear fractions prepared and immunoprecipitated with human sera²². The immune complexes were washed sequentially with buffer (50 mM Tris, pH 8.0, 10 mM EDTA, 0.05% NP40) containing 0.7 M, 0.6 M, and 0.5 M NaCl. **f**, Bacterial fusion protein (20 µg per lane), induced with 10 mM isopropyl-β-thiogalactopyranoside (IPTG) for 3 h at 37 °C, were immunoblotted with autoantibody. Autoantibody was affinity purified by elution with 1 ml of 3 M KSCN for 5 min at room temperature from nitrocellulose filters, lifted from confluent plates of IPTG-induced λ JPL4, or from an irrelevant phage. Eluates were diluted with 3 vol PBS, dialysed overnight against PBS at 4 °C and concentrated to 100 µg protein per ml. **g**, pJPZ1L4 (1 µg) was linearized and *in vitro* translated using the pBluescript T3 promoter (Promega). RNA transcripts were analysed in 1.0% agarose gels and *in vitro* translated using radiolabelled ^{35}S -methionine²⁶. The translated product was analysed by SDS-PAGE and immunoprecipitated with human sera.

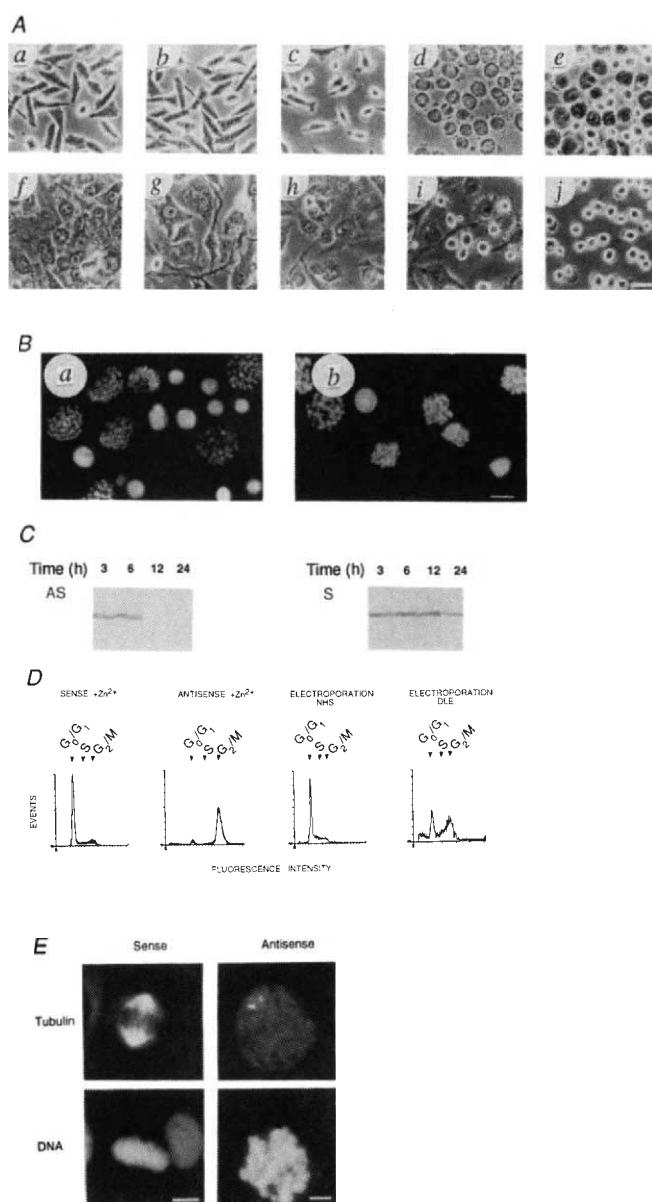


FIG. 4 Mitotic arrest and prevention of mitotic spindle formation in mouse L cells induced by antisense mRNA or by autoantibody. **A**, Progressive rounding and detachment of antisense-transfected cells (*b–e*) and of cells electroporated with affinity-purified autoantibody (*g–j*). Antisense-transfected L cells were treated with ZnCl₂ for 3 (*b*), 6 (*c*), 12 (*d*) and 24 h (*e*). **a**, Sense-transfected cells treated with ZnCl₂ for 24 h. L cells, electroporated with affinity-purified autoantibody, were examined at 3 (*g*), 6 (*h*), 12 (*i*) and 24 h (*j*). **f**, Cells electroporated with normal human IgG and examined at 24 h. Phase-contrast micrographs. Scale bar, 5 μ m. **B**, Mitotic arrest of L cells 24 h after induction of antisense mRNA by ZnCl₂ (*a*) or electroporation with affinity-purified autoantibody (*b*). Floating cells were collected, cytocentrifuged (700g for 5 min) and stained with 5 μ M Hoechst DNA dye. Scale bar, 2 μ m. **C**, Immunoblot of antisense transfected cells (AS) at 3, 6, 12 and 24 h after treatment with ZnCl₂ showing absence of RMSA-1 at 12 h; **S**, immunoblot of sense transfected cells treated with ZnCl₂; **D**, Flow cytometry showing cell-cycle arrest at G₂/M. Sense- and antisense-transfected L cells were synchronized by double thymidine block and treated with ZnCl₂. L cells were electroporated with affinity-purified autoantibody (DLE) or normal human IgG (NHS). Cells were sampled from 3 to 24 h. Only the 24-h time point is shown. **E**, Upper panels show immunofluorescence with anti-tubulin antibody; lower panels show the same cells stained with Hoechst DNA dye. Left-hand panels show sense- and right-hand panels antisense-transfected L cells cultured with ZnCl₂ for 24 h. Note absence of mitotic spindle in the antisense-transfected cell arrested in mitosis, whereas the sense-transfected cell shows a typical spindle. Scale bar: sense panels, 5 μ m; antisense panels, 2 μ m.

METHODS. pMT4SV_{neo} contains four copies of a synthetic mouse metallothionein promoter, G418 resistance gene and SV40 T-antigen gene^{13,14}. The T-antigen gene was replaced with full-length RMSA-1 cDNA in both orientations. L cells were transfected with plasmid DNA and selected with G418 (300 μ g ml⁻¹). Isolated colonies were picked and expanded into cell lines. Sense- and antisense-transfected cells were cultured with or without 100 μ M ZnCl₂. For electroporation, 5 $\times$ 10⁶ cells were subjected to 250 μ F at 0.45 kV constant voltage (BioRad). Cells were held in a cuvette with a path length of 0.8 cm holding 0.8 ml PBS, 10 mM MgCl₂ and 0.5 mg ml⁻¹ affinity-purified anti-RMSA-1 autoantibody or normal human IgG. After electroporation and 10 min incubation on ice, cells were washed, plated in culture medium and dead cells removed at 3 h by a change of medium. Antibody entry was confirmed by immunofluorescence and flow cytometry using FITC-labelled rabbit anti-human immunoglobulin. For immunofluorescence with anti-tubulin antibody, sense and anti-sense transfected cells were cultured on slides coated with poly-L-lysine (0.4 mg ml⁻¹) with or without 100 μ M ZnCl₂ for 24 h. The cells were fixed and permeabilized²¹, reacted with mouse monoclonal antibody to α -tubulin (Amersham) and traced with FITC-labelled anti-mouse immunoglobulin (Silenus). The cells were counterstained with Hoechst 33342 DNA dye.

during metaphase in the absence of chromosomes.

We suggest that RMSA-1 is activated by phosphorylation at the G₂/M transition by Cdc2 kinase because it is phosphorylated only in mitotic cells, contains consensus motifs for the kinase and is phosphorylated by the kinase *in vitro* (J.-P.Y., manuscript in preparation). Previous studies have suggested that spindle assembly is regulated by Cdc2 kinase^{1–5} and requires kinesin-related proteins^{16–18}, microtubule nucleation by centrosomes^{5,14,19}, microtubule capture and stabilization by kinetochores^{5,20}, and microtubule stabilization by mitotic chromatin^{5–8}, although kinetochores may not be obligatory^{5,6,21}. We suggest that RMSA-1 is a previously unidentified Cdc2 kinase substrate and chromatin factor which stabilizes spindle microtubules^{5,20}. □

Received 5 July; accepted 11 November 1993.

- Norbury, C. & Nurse, P. A. *Rev. Biochem.* **61**, 441–470 (1992).
- Murray, A. W. *Nature* **359**, 599–604 (1992).
- Lohka, M. J. & Maller, J. L. *J. Cell Biol.* **101**, 518–532 (1992).
- Verde, F., Labbe, J.-L., Doree, M. & Karsenti, E. *Nature* **343**, 233–238 (1990).
- Karsenti, E. *Semin. Cell Biol.* **2**, 251–260 (1991).
- Sawin, K. E. & Mitchinson, T. J. *J. Cell Biol.* **112**, 925–940 (1991).
- Karsenti, E., Newport, J., Hubble, R. & Kirschner, M. J. *Cell Biol.* **98**, 1730–1745 (1984).

- Nicklas, R. B. & Gordon, G. W. *J. Cell Biol.* **100**, 1–7 (1985).
- Shenoy, S. et al. *Cell* **57**, 763–774 (1989).
- Garcia-Bustos, J., Heitman, J. & Hall, M. N. *Biochim. biophys. Acta* **1071**, 83–101 (1991).
- Nigg, E. A. *Semin. Cell Biol.* **2**, 261–270 (1991).
- Peden, K. W. C., Charles, C., Sanders, L. & Tennekoon, G. I. *Expl Cell Res.* **185**, 60–72 (1989).
- Cicuttini, F. M. et al. *Blood* **80**, 102–112 (1992).
- Karsenti, E. & Maro, B. *Trends biochem. Sci.* **11**, 460–463 (1986).
- Picard, A., Harricane, M.-C., Labbe, J.-C. & Doree, M. *Devi Biol.* **128**, 121–128 (1988).
- Sanders, W. S. & Hoyt, M. A. *Cell* **70**, 451–458 (1992).
- Hagan, I. & Yanagida, M. *Nature* **356**, 74–76 (1992).
- Sawin, K. E., LeGuellec, K., Phillippe, M. & Mitchinson, T. J. *Nature* **359**, 540–543 (1992).
- Joshi, H. C., Palacios, M. J., McNamara, L. & Cleveland, D. W. *Nature* **356**, 80–82 (1992).
- Kirschner, M. & Mitchinson, T. J. *Cell* **45**, 329–342 (1986).
- Bernat, R. L., Borisy, G. G., Rothfield, N. F. & Earnshaw, W. C. *J. Cell Biol.* **11**, 1519–1533 (1990).
- Kooy, J., Toh, B. H., Pettit, J. M., Erlich, R. & Gleeson, P. A. *J. biol. Chem.* **267**, 20255–20263 (1992).
- Schickedanz, J., Scheidtmann, K. H. & Walter, G. *Virology* **148**, 47–57 (1986).
- Mu, F.-T. et al. *Clin. Exp. Immun.* **71**, 100–106 (1988).
- Bootsma, D., Budke, L. & Vos, O. *Expl Cell Res.* **33**, 301–309 (1964).
- Alderuccio, F., Chan, E. K. L. & Tan, E. M. *J. exp. Med.* **173**, 941–952 (1991).
- Young, R. A. & Davis, R. W. *Proc. natn. Acad. Sci. U.S.A.* **80**, 1194–1198 (1983).
- Hanahan, D. J. *molec. Biol.* **166**, 557–580 (1983).
- Chomczynski, P. & Sacchi, N. *Analyt. Biochem.* **162**, 156–159 (1987).
- Chen, E. Y. & Seeburg, P. H. *DNA* **4**, 165–170 (1985).

ACKNOWLEDGEMENTS. We thank E. Smith, P. Fergin and staff at Gribble's Pathology for blood samples, K. Peden for permission to use the pMT4 vector, F. Cicuttini for supplying the vector, K. Scarff for preliminary studies and I. R. van Driel for critically reading the manuscript. This work was supported by the Mary Whight Lupus Fellowship and the Anti-Cancer Council of Victoria.

Structure of the *Aeromonas* toxin proaerolysin in its water-soluble and membrane-channel states

Michael W. Parker*, J. Thomas Buckley†, Johan P. M. Postma‡, Alec D. Tucker‡§, Kevin Leonard‡, Franc Pattus‡ & Demetrius Tsernoglou‡

* St Vincent's Institute of Medical Research, 41 Victoria Parade, Fitzroy, Victoria 3065, Australia

† Department of Biochemistry and Microbiology, University of Victoria, British Columbia V8W 2Y2, Canada

‡ European Molecular Biology Laboratory, Meyerhofstrasse 1, D-6900 Heidelberg, Germany

§ Present address: ZENECA Pharmaceuticals, Alderley Park, Macclesfield, Cheshire SK10 4TG, UK

AEROLYSIN is chiefly responsible for the pathogenicity of *Aeromonas hydrophila*, a bacterium associated with diarrhoeal diseases and deep wound infections¹. Like many other microbial toxins, the protein changes in a multistep process from a completely water-soluble form to produce a transmembrane channel that destroys sensitive cells by breaking their permeability barriers². Here we describe the structure of proaerolysin determined by X-ray crystallography at 2.8 Å resolution. The protoxin (*M*, 52,000) adopts a novel protein fold. Images of an aerolysin oligomer derived from electron microscopy have assisted in constructing a model of the membrane channel and have led to the proposal of a scheme to

account for insertion of the protein into lipid bilayers to form ion channels.

The proaerolysin monomer is conveniently divided into four domains (Fig. 1a). Domain 1 is a small lobe protruding from a larger lobe containing the other three domains. The larger lobe has an unusual elongated shape, 110 Å long and 20–45 Å in thickness. The protein is composed of 40% β -sheet structure and 17% α -helix. The β -structure consists of twenty β -strands that take part in eight antiparallel β -sheets. Two strands (residues 169–206 and 290–322) span the major lobe of the molecule almost tip to tip. One of them measures ~93 Å, far longer than the longest strand in the current release of the Brookhaven Data Base³ (a 48 Å-long stretch of the coat protein of rhinovirus⁴). In passing from one end of the large lobe to the other, the strands adopt a helical twist so that the hydrophilic face of the sheet in domain 2 undergoes a 180° rotation as it reaches domain 4. Analysis of the distribution of side chains in the structure indicates significant clustering of aromatic residues on the surface of domain 2 and in the core of domain 3 (Fig. 1b). Also, about 70% of the aromatic residues in the molecule are at least partially accessible to solvent.

The two monomers in the asymmetric unit of the crystal are related by an almost exact twofold axis (r.m.s. deviation between α -carbon atomic positions on superposition is 0.86 Å; Fig. 2a, b). Domain 1 is stabilized in its position by contacts with domain 1 of the other monomer. A striking feature of the interaction is that the domains clasp each other in a crossover fashion (Fig. 2b). The monomer-monomer interface is fairly extensive, with ~2,500 Å² buried per monomer. There are five salt bridges and twenty residues involved in hydrogen-bonding interactions between the monomers, which are evenly distributed between the domains. There is a non-uniform distribution of polar and non-polar residues in the interfacial regions (Fig. 2c, d), however, unlike that normally found in protein subunit interfaces⁵.

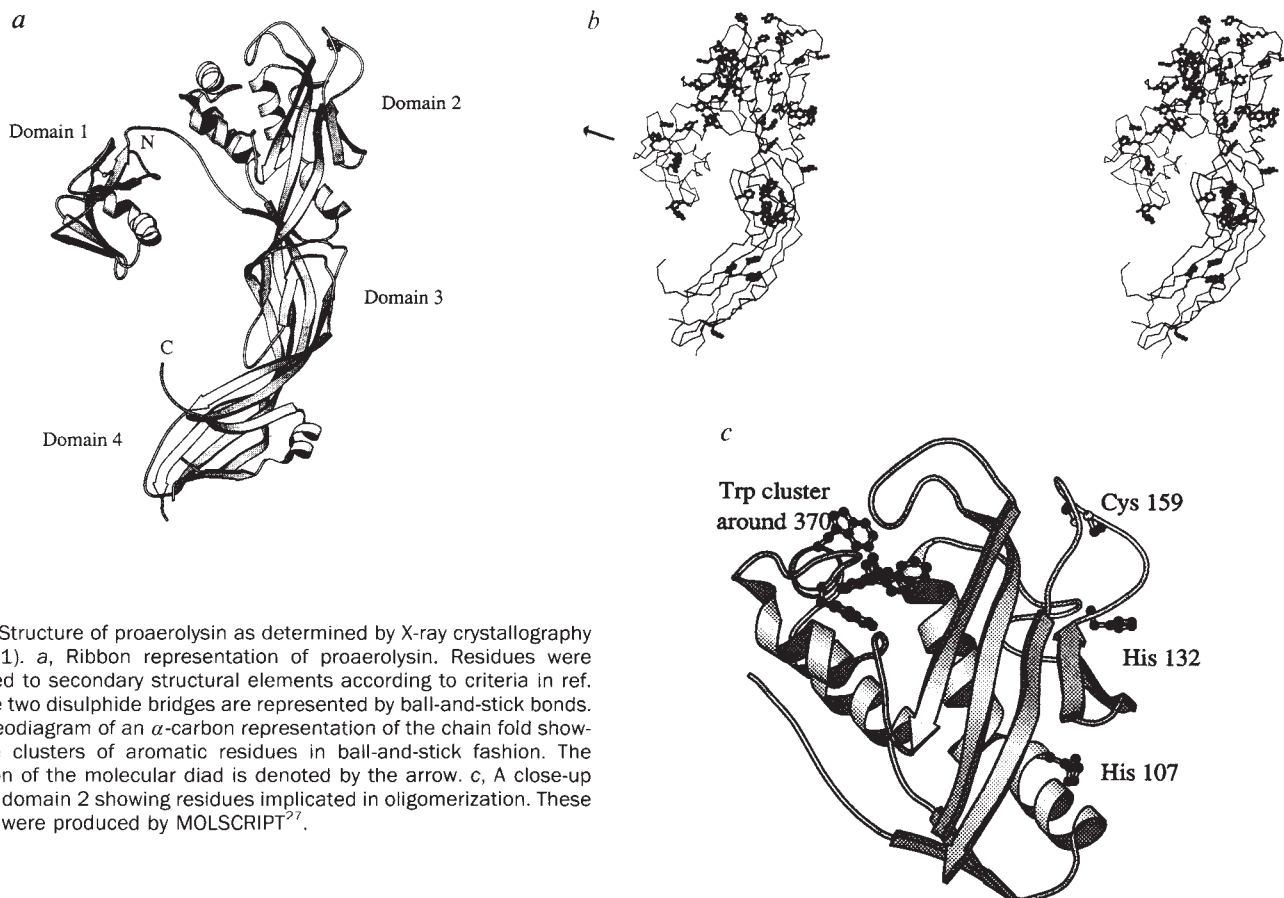


FIG. 1 Structure of proaerolysin as determined by X-ray crystallography (Table 1). a, Ribbon representation of proaerolysin. Residues were assigned to secondary structural elements according to criteria in ref. 26. The two disulphide bridges are represented by ball-and-stick bonds. b, Stereodigraph of an α -carbon representation of the chain fold showing the clusters of aromatic residues in ball-and-stick fashion. The direction of the molecular diad is denoted by the arrow. c, A close-up view of domain 2 showing residues implicated in oligomerization. These figures were produced by MOLSCRIPT²⁷.

TABLE 1 Summary of diffraction data and phasing statistics

	Native	KAu(CN) ₂	K ₂ HgI ₄	Pb(CH ₃ COO) ₂	K ₃ UO ₂ F ₅
Soak time (d)	—	1	1	4	1
Soaking Concentrations (mM)	—	11.5	0.5	10.0	0.5
Resolution (Å)	2.8	3.0	3.0	3.2	3.1
No. of observations	145,443	126,188	61,019	95,468	51,892
No. of unique reflections	30,060	24,525	23,124	20,320	21,162
Data completeness (%)	98	98	93	98	93
No. of data >3σ(%)	79	82	81	76	75
R _{merge} * (%)	9.1	9.9	9.7	13.3	12.9
MFID † (%)	—	15.5	20.9	12.4	34.2
No. of sites	—	5	4	4	3
R _{Cullis} ‡ (%)	—	57.7	61.7	59.2	60.4
< F _H >/E [§]	—	1.7	0.6	1.3	1.0

Proaerolysin from *A. hydrophila* was overproduced, purified and crystallized as before¹⁹. The crystals belong to space group P4₃2₁2, with cell dimensions $a = b = 104.0$ Å, $c = 222.0$ Å. There are two monomers in the asymmetric unit. Data collection: Data used in the structure determination were initially collected to low resolution (6.5 Å) on an in-house multiwire area detector (Nicolet) and later recollected on an image plate detector on the X-31 beamline at the Deutsche Elektronen Synchrotron, EMBL Outstation, Hamburg. The latter data sets were processed using programs in the CCP4 suite²⁰. Heavy-atom derivatives: Crystals were soaked in artificial mother liquor containing the derivative at room temperature. In the case of the uranyl derivative, the concentration of the heavy atom was slowly increased over several hours to the final concentration. Each dataset was collected from one crystal. Two major sites were located in the isomorphous difference Patterson function of the uranyl derivative and subsequent sites were found by cross-difference Fourier analysis. Anomalous scattering data were collected for all four derivatives and were used to establish the correct enantiomorphic space group unequivocally. Heavy-atom parameters were refined against centric data and phasing was carried out using the CCP4 programs REFIN and PHASE²⁰. The overall figure-of-merit was 0.52 (for the resolution shell 15.0–3.0 Å). Solvent flattening: The initial MIR map was uninterpretable but could be improved with iterative cycles of solvent flattening²⁰. A solvent density of 45%, much lower than the estimated content of about 58% from the cell parameter data, was imposed to ensure all the electron density was included in the protein mask. The starting model was built into the solvent-flattened electron density map on an Evans and Sutherland PS390 graphics system using FRODO²¹. Density modification: A starting envelope for density averaging was created from an early model and twofold averaging carried out using the DEMON program suite²². Refinement: Partial models were improved with molecular dynamics refinement using GROMOS²³. Phases from these models were combined with MIR and solvent-flattened phases using COMBINE²⁰. In all, six rounds of model building and MD refinement were required to produce a penultimate model which was subsequently refined using PROLSQ²⁴. Most model rebuilding and analysis was done on an Evans and Sutherland ES330 graphics workstation using FRODO²¹ and MIDAS²⁵. Model quality: The map interpretation was checked using residue omit maps. About 10% of the model was deleted, a few cycles of refinement performed to reduce bias, and the resultant $(2|F_{\text{obs}}| - |F_{\text{calc}}|)$ electron density map examined in the region of omission. The correctness of the tracing is supported by the convincing fit of the 56 aromatic residues found in proaerolysin into the electron density map. Furthermore, all 52 heavy-atom sites found during the course of solving the structure are located in chemically reasonable positions in the model. The current model of proaerolysin includes residues 2–207, 216–422, and 440–470 in monomer A, and 1–207, 213–424, and 441–468 in monomer B. The missing residues lie in loop regions of disordered structure. There are no solvent molecules included at this stage. The crystallographic *R*-factor is 23.0% for all measurements between 6–2.8 Å. The r.m.s. deviations from ideal geometry of the model are 0.019 Å and 2.9° for bond angles. Tightly restrained individual isotropic temperature factors (r.m.s. deviation in *B*-factor between bonded main-chain atoms of 1.0 Å²) were employed in the latter stages of refinement. The path of the polypeptide is clear and most of the side chain density is well defined, with the exception of a few charged surface residues. The proaerolysin coordinates will be deposited in the Brookhaven Protein Data Bank. Further details of the refined structure will be published elsewhere.

* $R_{\text{merge}} = \sum_i |I_i - \langle I \rangle| / \sum_i \langle I \rangle$, where I_i is the intensity for the i th measurement of an equivalent reflection with indices h, k, l .

† MFID = $\sum_i |F_{\text{PH}}| - |F_{\text{P}}| / \sum_i |F_{\text{P}}|$ where F_{PH} refers to derivative data and F_{P} to native data.

‡ $R_{\text{Cullis}} = \sum_i |F_{\text{PHC}}| - |F_{\text{PH}}| / \sum_i |F_{\text{PH}}| - |F_{\text{P}}|$. The summation is over all centric reflections. F_{PHC} and F_{PH} are calculated and measured derivative structure factor amplitudes respectively. F_{P} is the native structure factor.

§ $\langle |F_{\text{H}}| \rangle / E$ is a measure of the phasing power of the derivative, where $\langle |F_{\text{H}}| \rangle$ is the root-mean-square heavy-atom derivative structure factor amplitude and E is the lack-of-closure error.

Recently we discovered that proaerolysin is a dimer in solution as well as in the crystal⁶. Interactions between extensive patches of hydrophobic residues observed in the monomer–monomer interface (Fig. 2c, d) may provide a significant driving force in the association of the monomers.

Proaerolysin is activated outside the cell by proteolytic nicking⁷, for example trypsin cuts after Lys 427 (ref. 8). The cleavage region is not observed in the crystal structure, but its location can be inferred to be in a long, flexible loop at the tip of domain 4, based on the position of neighbouring residues (Fig. 1a). The sequence distal to the site of processing (to residue 470), covers ~1,100 Å² in domain 4 (Fig. 2e). Dissociation of the distal peptide(s) would expose a large hydrophobic patch (Fig. 2f), but proteolytic cleavage is probably not sufficient by itself to dislodge the peptide(s) because of the extensive hydrophobic contacts, and the two salt bridges and eighteen hydrogen bonds formed with the rest of domain 4. The processed toxin remains a dimer and appears to have a very similar structure to the unprocessed form, as judged by circular dichroism⁶.

Oligomerization is an essential step in channel formation^{7,9,11} and it seems to precede membrane insertion^{7,11,12}. A variety of factors promote oligomerization, including low ionic strength, high protein concentration and the presence of ionic detergents⁸.

The binding of a hydrophobic dye to oligomers, but not to processed dimers, suggests that the activation peptide is released upon oligomerization¹¹. Oligomers can rapidly aggregate and require ionic detergents to stay monodisperse¹¹. Mutagenesis studies have implicated His 107, His 132, Cys 159, Trp 371 and Trp 373, which are all located in domain 2, in oligomerization^{9,11,12} (Fig. 1c), suggesting that domain 2 is important in this step.

Image analysis from electron micrographs of two-dimensional crystals of aerolysin in lipid membranes has yielded a low-resolution model of the aerolysin channel¹³ (Fig. 3a, b). We have attempted to fit the water-soluble structure of aerolysin (minus activation peptide) into this image (Fig. 3c, d). The fit is based on the assumption, supported by circular dichroism measurements on preformed oligomers (results not shown), that there are no major structural changes to the individual domains. Because domains 2, 3 and 4 share a number of β-strands, the large lobe of proaerolysin has been treated as one rigid body. A convincing fit to the image can be obtained if domain 1 is allowed to rotate relative to the larger lobe. As mentioned earlier, this domain is only held in position by virtue of the dimer state: cleavage of the dimer to monomers would free it to rotate and bury exposed hydrophobic regions. There are no serious clashes in the model

FIG. 2 Space-filling representations of proaerolysin generated by MIDAS²⁵ on a Silicon Graphics Workstation 4D/480VGX. *a, b*, Two views of the proaerolysin dimer. Each monomer is indicated by a different colour. *c*, The proaerolysin monomer viewed down the monomer-monomer interface that covers the region of domains 2–4. Hydrophobic and aromatic residues are coloured in green. *d*, As *c*, but viewed down the monomer-monomer interface that covers the region of domain 1. *e*, Proaerolysin monomer with the activation peptide coloured blue. *f*, Aerolysin monomer minus the activation peptide. Hydrophobic and aromatic residues are coloured green.

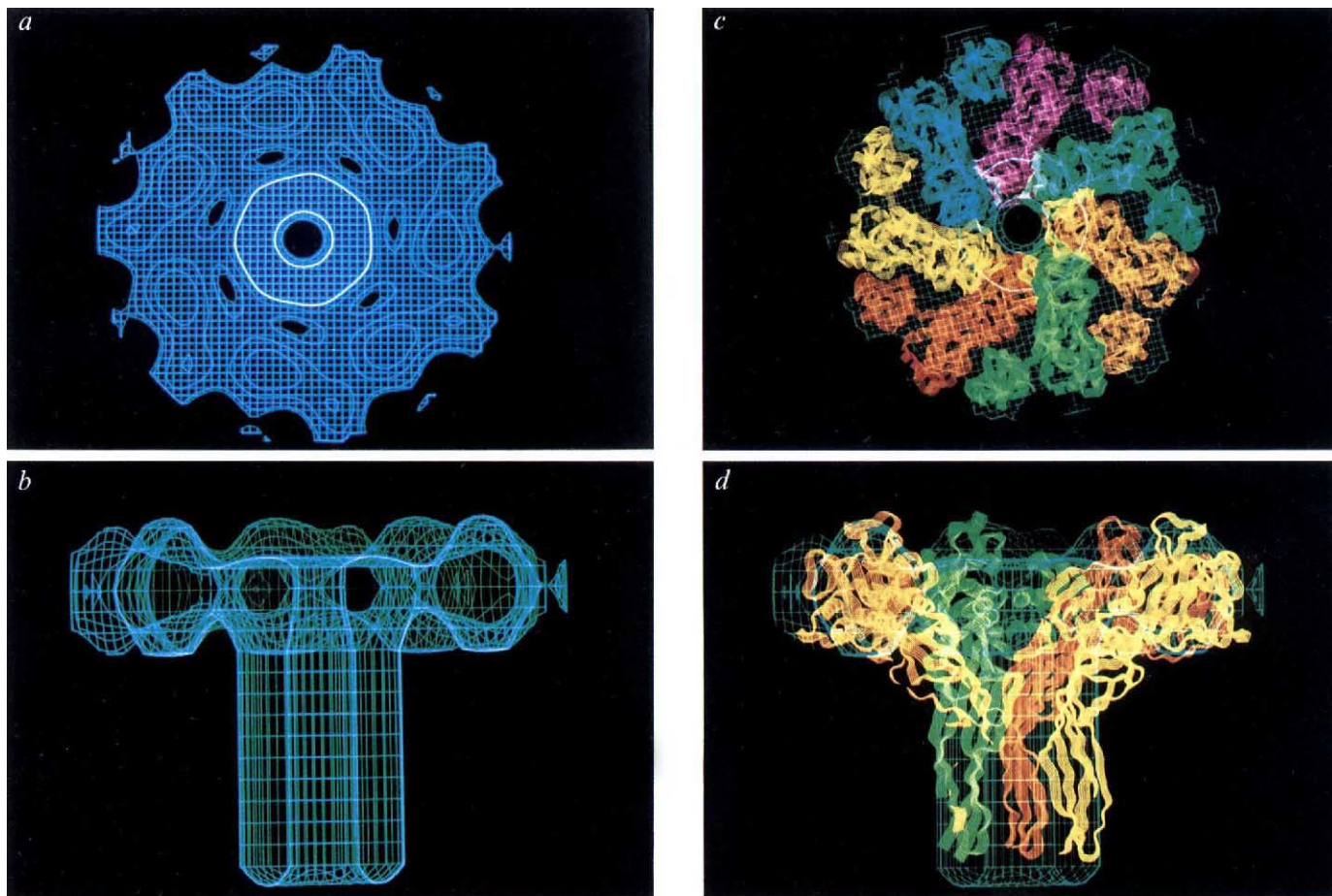
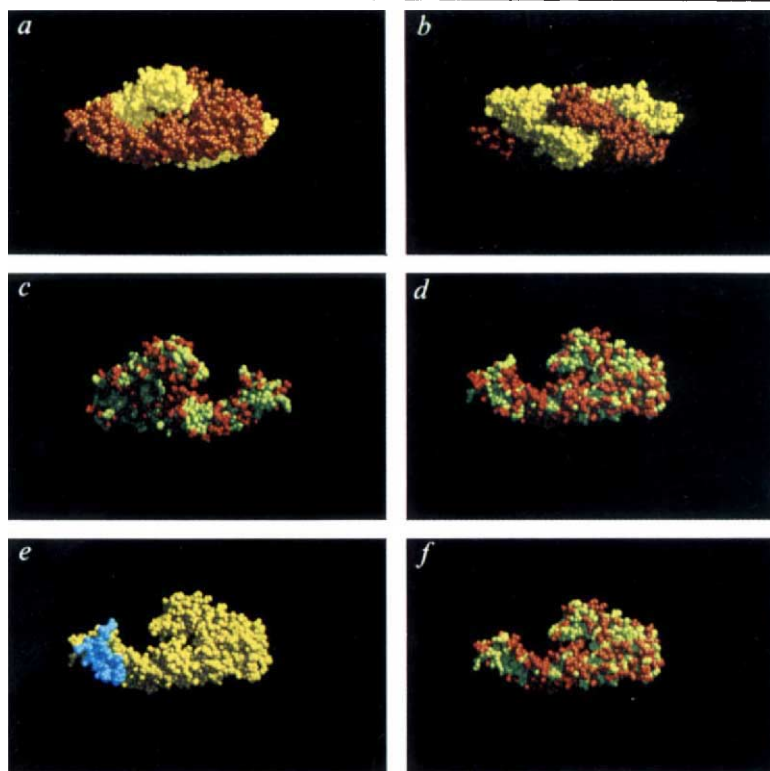


FIG. 3 *a*, Image of the aerolysin channel derived from electron microscopy when viewed down onto the membrane surface. The resolution of the image is ~ 25 Å. The image consists of a central cylindrical-shaped density of outer diameter ~ 46 Å encircling a water-filled channel 17 Å in diameter. The cylinder is surrounded by arms of density, each of which are made up of two unequal-sized domains. The total diameter of the disk-like complex is ~ 140 Å. (We estimate the uncertainty in

these measurements to be $\sim 5\%$.) *b*, Side view of the image derived from electron microscopy. The disk is 30–40 Å thick and lies above the membrane surface at a distance of ~ 20 Å. The cylindrical-shaped density projecting from the disk is 50 Å in length, of which the last 30 Å spans the lipid bilayer. *c, d*, Fit of the aerolysin monomers into the channel density. Domain 4 of each aerolysin monomer is traversing the lipid bilayer.

between monomers, and the hydrophobic surfaces that were buried in the proaerolysin dimer interface are now buried in the monomer interfaces of the oligomer. This model explains the sensitivity of domain 2 mutants towards oligomerization and predicts that mutations in domains 1 and 4 will exhibit similar effects. In this model, domain 4 spans the lipid bilayer with three β -strands per monomer lining the channel lumen. The hydrophobic patch revealed by removing the activation peptide faces the lipid core of the membrane.

The models of the water-soluble dimer and the membrane-bound oligomer suggest a pathway for the biogenesis of the aerolysin channel. The protoxin would approach the target cell as water-soluble dimer and be concentrated on the surface by binding to its receptor. Proteolytic nicking at the activation site would lead to oligomerization, requiring dissociation of the dimer and removal of the activation peptide. These events might be promoted by the detergent properties of the nearby lipid bilayer. The resulting exposure of hydrophobic regions of the

toxin would result in an energetically favourable state for membrane insertion. The oligomer would partition into the lipid bilayer from the domain 4 end because it would still have an exposed hydrophobic face, produced by removal of the activation peptide, which could not be buried in the oligomerization step. This end of the oligomer, composed almost entirely of β -structure, would adopt a barrel topology similar to that found in porins^{14,15}, resulting in a transmembrane pore.

Transition from a water-soluble to amphipathic state is an essential step in a variety of biological processes, including membrane penetration by many bacterial toxins², complement action¹⁶, pH-dependent transition of viral fusion proteins¹⁷ and post-translational insertion of proteins into the membranes of various organelles¹⁸. All these events require the exposure of hydrophobic regions immediately preceding membrane penetration. The structure of proaerolysin suggests that multimer dissociation and peptide removal are two means by which this may occur. □

Received 28 July; accepted 25 November 1993.

1. Altmegg, M. & Geiss, H. K. *CRC Crit. Rev. Microbiol.* **16**, 253–286 (1989).
2. Parker, M. W., Tucker, A. D., Tsernoglou, D. & Pattus, F. *Trends biochem. Sci.* **15**, 126–129 (1990).
3. Bernstein, F. C. et al. *J. molec. Biol.* **112**, 535–542 (1977).
4. Arnold, E. & Rossmann, M. G. *J. molec. Biol.* **211**, 763–801 (1990).
5. Argos, P. *Protein Eng.* **2**, 101–113 (1988).
6. van der Goot, F. G., Ausio, J., Wong, K. R., Pattus, F. & Buckley, J. T. *J. biol. Chem.*, **268**, 18272–18279 (1993).
7. Garland, W. J. & Buckley, J. T. *Infect. Immunity* **56**, 1249–1253 (1988).
8. van der Goot, F. G. et al. *Biochemistry*, **31**, 8566–8570 (1992).
9. Green, M. J. & Buckley, J. T. *Biochemistry* **29**, 2177–2180 (1990).
10. Wilmsen, H.-U., Pattus, F. & Buckley, J. T. *J. Membrane Biol.* **115**, 71–81 (1990).
11. van der Goot, F. G., Pattus, F., Wong, K. R. & Buckley, J. T. *Biochemistry* **32**, 2636–2642 (1993).
12. Wilmsen, H.-U., Buckley, J. T. & Pattus, F. *Molec. Microbiol.* **5**, 2745–2751 (1991).
13. Wilmsen, H.-U., Leonard, K. R., Tichelaar, W., Buckley, J. T. & Pattus, F. *EMBO J.* **11**, 2457–2463 (1992).
14. Weiss, M. S. et al. *Science* **254**, 1627–1630 (1991).
15. Cowan, S. W. et al. *Nature* **358**, 727–733 (1992).

16. Bhakdi, S. & Tranum-Jensen, J. *Rev. Physiol. Biochem. Pharmac.* **107**, 148–156 (1987).
17. Wiley, D. C. & Skehel, J. J. *Rev. Biochem. Sci.* **56**, 365–394 (1987).
18. Wickner, W. *Trends biochem. Sci.* **14**, 280–283 (1989).
19. Tucker, A. D., Parker, M. W., Tsernoglou, D. & Buckley, J. T. *J. molec. Biol.* **212**, 561–562 (1990).
20. CCP4 (Daresbury Laboratory, Warrington WA4 4AD, UK, 1979).
21. Jones, A. J. *Appl. Crystallogr.* **11**, 268–274 (1978).
22. Vellieux, F. M. D. et al. *Proc. natn. Acad. Sci. U.S.A.* **90**, 2355–2359.
23. van Gunsteren, W. F. *Protein Engng* **2**, 5–13 (1988).
24. Hendrickson, W. A. & Konnert, J. M. in *Computing in Crystallography* (eds Diamond, R., Ramaseshan, S. & Venkatesan, K.) 13.01–13.23 (Indian Academy of Science, Int. Union of Crystallography, Bangalore, 1980).
25. Ferrin, T. E., Huang, C. C., Jarvis, L. E. & Langridge, R. J. *molec. Graph.* **6**, 13–27 (1988).
26. Kabsch, W. & Sander, C. *Biopolymers* **22**, 2577–2637 (1983).
27. Kraulis, J. P. *J. appl. Crystallogr.* **24**, 946–950 (1991).

ACKNOWLEDGEMENTS. We thank G. van der Goot for discussions. M.W.P. is a Wellcome Australian Senior Research Fellow supported by the Australian Research Council and the National Health and Medical Research Council. J.T.B. is supported by the National Science and Engineering Council of Canada.

Direct interaction of human TFIID with the HIV-1 transactivator Tat

Fatah Kashanchi, Graziella Piras, Michael F. Radonovich, Janet F. Duvall, Ali Fattaey*, Cheng-Ming Chiang†, Robert G. Roeder† & John N. Brady

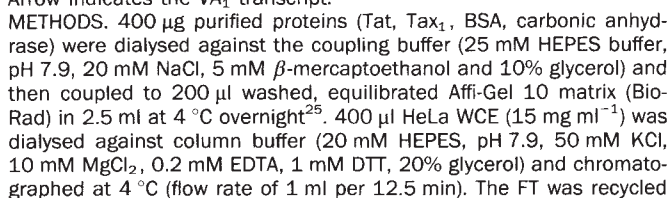
Laboratory of Molecular Virology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA
* Massachusetts General Hospital Cancer Center, Building 149, 13th Street, Charlestown, Massachusetts 02129, USA
† The Rockefeller University, 1230 York Avenue, New York, New York 10021, USA

THE *tat* gene of the human immunodeficiency virus (HIV) plays a central role in the activation and life cycle of HIV. The *tat* protein (Tat) specifically transactivates HIV transcription *in vivo* and *in vitro*^{1–8}, exerting its effects at the level of transcriptional initiation and elongation. Here we report that Tat binds directly to the basal transcription factor TFIID. The transcriptional activity of HeLa extracts was depleted after chromatography on a Tat affinity column, which specifically retained the polymerase II-specific factor TFIID. Direct interaction of Tat with holo-TFIID, composed of TATA-binding protein (TBP) and associated factors (TAFs), was observed. Tat binds, through amino acids 36–50, directly to the

TBP subunit of TFIID. Our results suggest that Tat may transduce upstream or downstream regulatory signals by direct interaction with the basal transcription factor TFIID.

Transcriptionally active HeLa whole cell extract was applied to a Tat or control protein affinity columns. The flow-through was analysed for basal transcription activity on the adenovirus major late (AdML) and HIV promoters. Whereas the flow-through from the control columns could support transcription of the AdML and HIV templates, that from the Tat affinity column was inactive (Fig. 1a; compare lanes 3 and 4 with 1, 2 and 5–10; Fig. 1c, lane 2 versus 5 and 6). Transcriptional activity was reconstituted with the addition of recombinant human TBP (rhTBP) (Fig. 1b, lane 2; Fig. 1c, lane 3), but not rhTFIIIB (Fig. 1b, lane 3; Fig. 1c, lane 4) to the depleted flow-through. Proteins were eluted from the Tat and control affinity columns, concentrated and added to TFIID heat-inactivated HeLa extracts (Fig. 1d, lanes 1 and 2) (ref. 9). Addition of rhTBP, but not rhTFIIIB, to the extract restores transcriptional activity (Fig. 1d, lane 7; data not shown). The 1 M KCl elution fraction from the Tat, but not the control, affinity column specifically reconstituted HIV transcription (Fig. 1d, lanes 4 and 6).

TBP-TAF complexes are heterogeneous and involved in RNA polymerase (pol) I, II (TFIID) and III (TFIIIB) transcription^{10–15}. Tat does not interact with all species of TBP-containing complexes, for example TFIIB (Fig. 1e). In addition, in the Tat affinity column flow-through fraction, TFIID binding is substantially decreased, but not totally depleted (Fig. 2a and b). Thus, even though Tat is in excess, not all TFIID is bound to the Tat affinity column. The TFIID that does not associate with Tat may not support, or may not be in sufficient concentra-



NATURE · VOL 367 · 20 JANUARY 1994

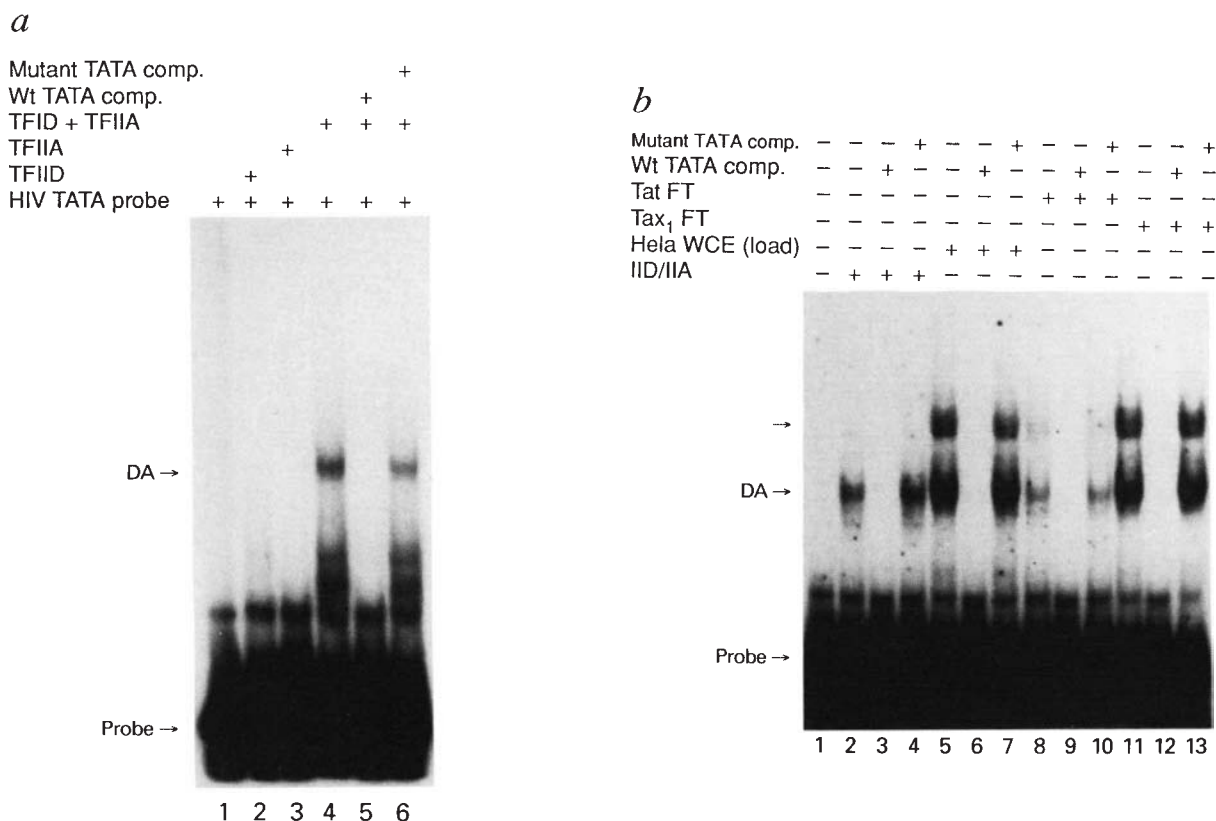


FIG. 2 Gel shift assay using partially purified transcription factors TFIID and TFIIA. **a**, HeLa transcription factors were purified as described²⁶. Lane 1, probe; lane 2, HeLa TFIID (DEAE; 1 μ l); lane 3, HeLa TFIIA (Mono-S; 1 μ l); lane 4, TFIID and TFIIA; lanes 5 and 6, competition (comp.) in the presence of either unlabelled TATA oligonucleotide (lane 5, 100 ng) or the 5-base mutant TATA oligonucleotide (lane 6, 100 ng). **b**, Gel shift assay using affinity column FT. Lanes 1–5 are described in **a**; lanes 5–7, HeLa WCE (15 μ g load); lanes 8–13, FT from either the Tat or Tax₁ affinity columns (15 μ g).

METHODS. The probe (6.25 ng per reaction) used was the HIV promoter sequence –38 to –8 (5'-GATGCTGCATATAAGCAGCTGCTTTTGCC-3'). The mutant oligonucleotide was identical except that the underlined 5'-TATAA-3' sequence was changed to 5'-ACACT-3'. Two gel shift complexes were observed with the WCE, one representing the TFIID/TFIIA (DA) complex and a slower migrating complex (arrow) that probably represents the association of TFIIB, TFIIF and polymerase II into a higher order complex. Competition assays were as for **a**. Gel shift reactions and electrophoresis have been described²⁷.

tion to support, efficient basal transcription. Thus, even though TFIID is present in the flow-through fraction, no transcriptional activity is observed. Consistent with these results, in coprecipitation analysis Tat appears to interact with a select species of pol II TFIID (data not shown).

Interaction of Tat with a glutathione *S*-transferase (GST) fusion protein GST-TBP, but not with GST or a GST-Rb fusion protein, was observed in binding assays *in vitro* (Fig. 3a). Similar studies with the GST-TFIIB protein failed to demonstrate specific interaction of Tat with TFIIB (data not shown). Peptide competition analysis demonstrated that Tat-TBP binding was mediated through the conserved core domain of Tat. Peptide 36–50 specifically and reproducibly competed for binding to GST-TBP (72%) (Fig. 3b). In several independent experiments, we consistently observed a 70–80% reduction in GST-TBP/Tat binding in the presence of the 36–50 peptide. Less than a 20% reduction was seen with peptides covering other domains of the Tat protein. Mutation of the highly conserved amino acid Lys 41 has been shown to decrease the level of transactivation *in vivo*^{16–18}. In contrast to the wild-type peptide, a Lys 41 mutant peptide (Lys 41 to Thr 41) failed to inhibit Tat interaction with GST-TBP (Fig. 3c) or to bind to GST-TBP directly (Fig. 3d).

We next analysed the ability of a full-length 86-amino-acid Lys 41 Tat mutant protein to retain TFIID on an affinity column. The activity of the Lys 41 Tat mutant was first compared with wild-type Tat in two control assays, transactivation

response element (TAR) RNA gel shift and transactivation. The mutant was active for TAR RNA binding, but transcriptionally inactive (Fig. 4a and b). HeLa whole cell extract was chromatographed over Tat, Tat-Lys 41 mutant, or control affinity columns, and the flow-through assayed for basal transcription. Transcription was observed with the flow-through from the Tat mutant column but not the Tat wild-type column (Fig. 4c, lanes 4 and 5), demonstrating that the Tat mutant did not interact with TFIID.

Holo-TFIID is a complex of TBP and associated TAFs, including p250 (refs 19–22). Tat was incubated with DEAE-purified holo-TFIID and precipitated with anti-Tat serum. TAF p250 was present in the immunoprecipitate, demonstrating that Tat interacts with holo-TFIID (Fig. 4d). This interaction could be blocked by a wild-type, but not the Lys 41 mutant, Tat peptide 36–50. In a separate series of experiments, we have demonstrated that the species of holo-TFIID interacting with Tat also contains TAF p55 (data not shown). We have also used the epitope-tagged TBP (eTBP) HeLa cell line, LTR α , to purify holo-TFIID. The TAFs were kinase-labelled with ³²P (ref. 19), the holo-TFIID complex was affinity-purified and then incubated with GST-Tat. These experiments show that p250, p125, p70, eTBP and p30 are present in the Tat-holo-TFIID complex (data not shown). The presence of other TAFs is being investigated.

We have demonstrated a direct and specific interaction

between Tat and TFIID. A single amino-acid substitution at Lys 41 abolishes this protein-protein interaction. Lysine 41 is critical for transactivation *in vivo*, but does not decrease TAR RNA binding, suggesting that the Tat-TBP interaction is mechanistically related to transcriptional activation. We have observed that Tat increases the interaction of TFIID with the

HIV promoter (data not shown). Whether Tat increases the association or decreases dissociation of TFIID, or recruits a particular species of the functionally heterogeneous TFIID to the HIV template, is under investigation. Our results do not rule out the possibility that Tat may first bind to TAR RNA, then interact selectively with TFIID in the initiation complex to facili-

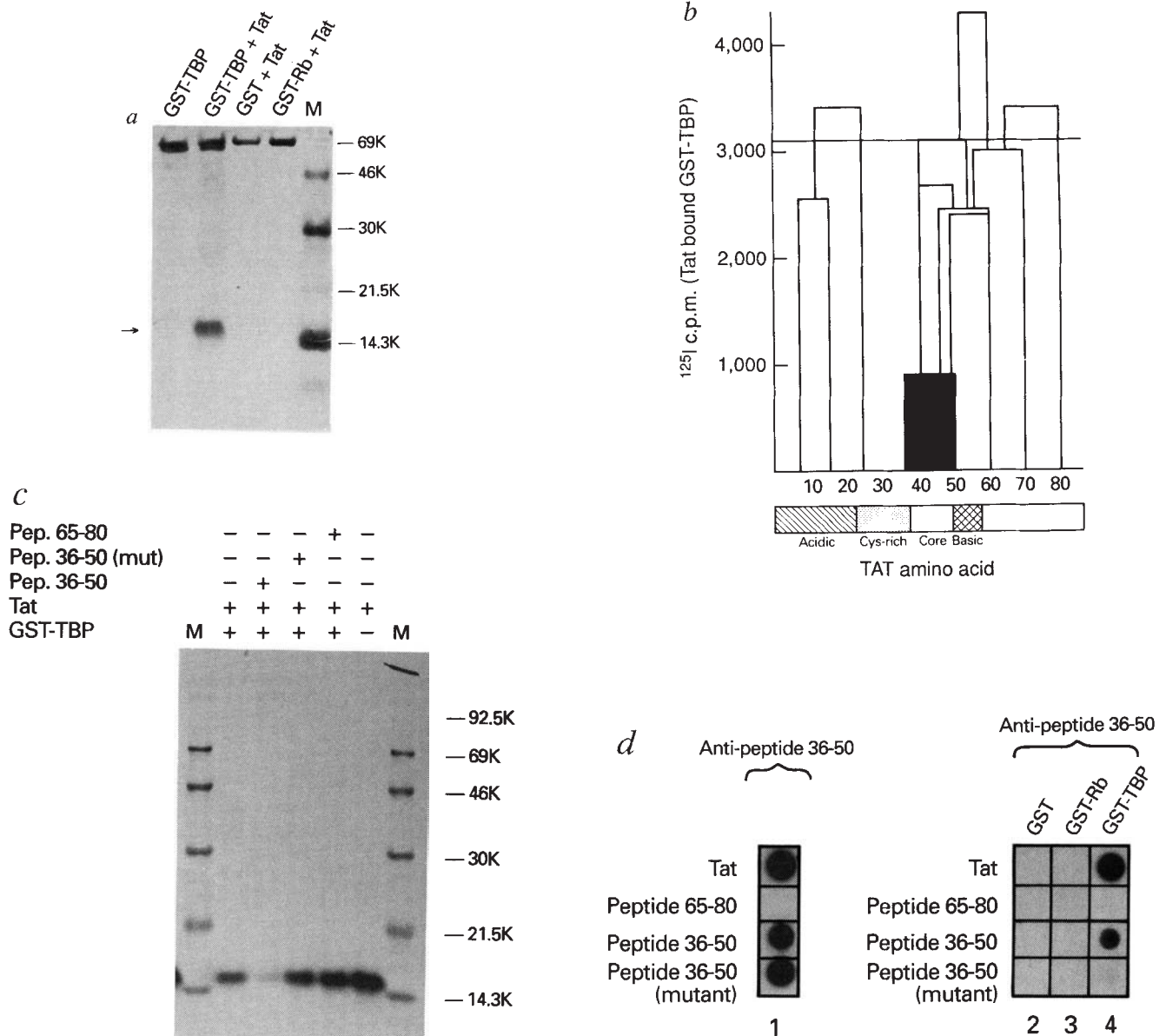
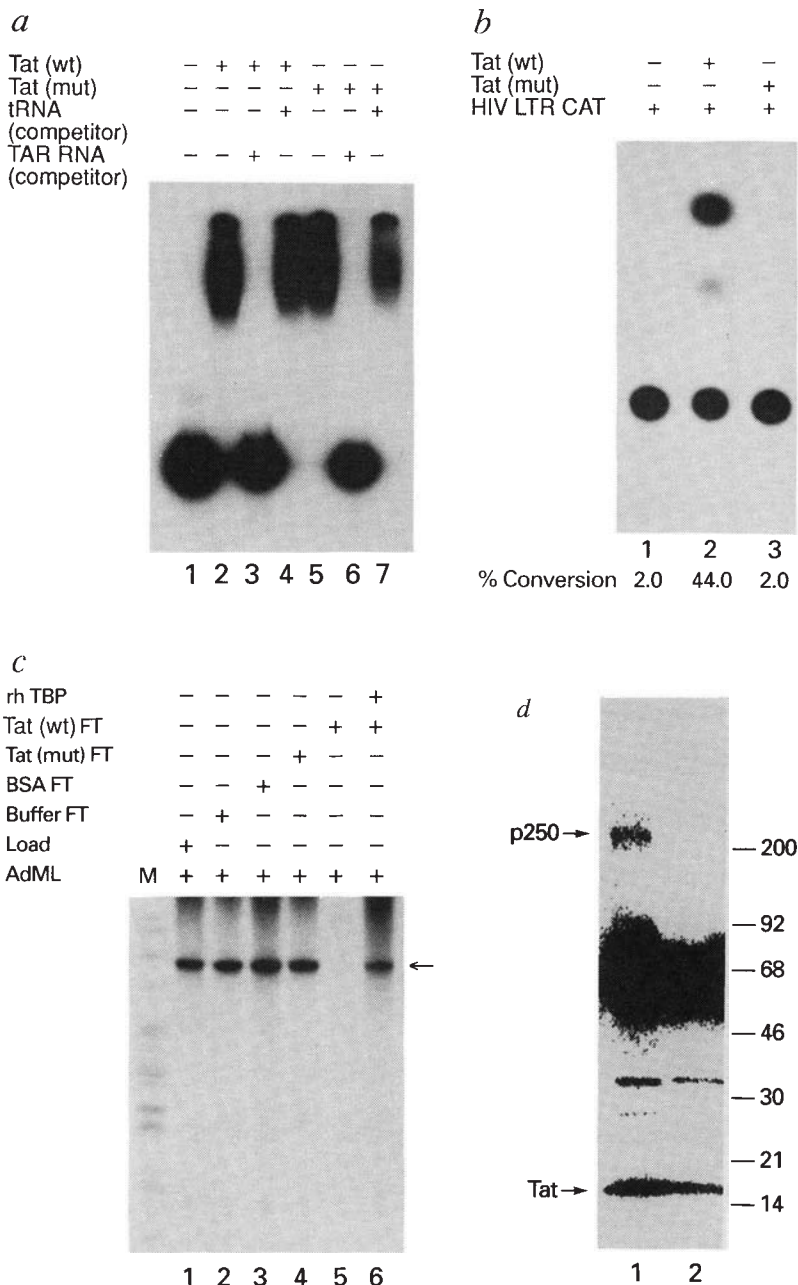


FIG. 3 Binding of Tat to glutathione S-transferase-TBP fusion protein. **a**, Binding of purified Tat protein to GST or GST fusion proteins. Arrow represents the 17K Tat protein. **b**, Ten overlapping peptides were pre-incubated with GST-TBP, then wild-type Tat was added. Line extending across the bar graph represents the level of Tat binding in the absence of competitor peptide. **c**, As **b**, except that a mutant peptide of 36-50 (Lys 41→Thr) was added as a competitor (lane 3). **d**, Binding of Tat and Tat peptide 36-50 to GST-TBP. Lane 1 represents the ability of a polyclonal antiserum against peptide 36-50 to recognize wild-type Tat, peptide 36-50 (wild-type and mutant), but not peptide 65-80. METHODS. GST fusion proteins were purified as described²⁸. All fusion proteins were confirmed by western blotting with appropriate antibodies. **a**, Purified Tat protein (~11 µg) and GST or GST fusion proteins (GST-TBP, GST-Rb; 7.5 µg each) were mixed and incubated on ice for 1 h in the presence of buffer D and 0.15% NP-40 (reaction volume 20 µl). After incubation, 30 µl glutathione-agarose beads (packed volume, washed 3 times in buffer D and 0.15% NP-40) were added. Bound proteins were

denatured and separated by 14% SDS-PAGE. Gels were either stained with Coomassie blue or transferred for western blotting with Tat-specific antiserum. **b**, Ten overlapping peptides, 20 µg each in Tat buffer, were incubated with GST-TBP (7.5 µg) on ice for 30 min. Wild-type Tat (~11 µg) was then added to the mixture and incubated for 30 min on ice, followed by glutathione beads and incubation for a further 30 min. Unbound proteins were removed by washing (3 times, 1 ml each) in buffer D with 0.15% NP-40. Bound proteins were denatured and run on 14% SDS-PAGE and analysed by western blotting using antiserum and ¹²⁵I-labelled protein A. Tat-specific bands were cut from the blot and ¹²⁵I-protein A counted. **d**, Western dot blot of Tat protein and Tat peptides. GST-TBP, GST-Rb or GST proteins (7.5 µg) were mixed with Tat protein (5 µg) or various Tat peptides (30 µg) and the binding assayed. Pelleted beads were resuspended in 'stop' buffer (150 mM NaCl, 20 mM Tris, pH 7.5, 1% SDS), vortexed and centrifuged. Supernatants were dotted onto a nitrocellulose membrane (S&S, PH79, 0.1 µm pore size), air-dried for 30 min and blocked for 2 h in 5% non-fat milk.

FIG. 4 Analysis of wild-type and the Lys-41 Tat mutant by TAR RNA gel shift, transactivation and TBP binding. **a**, 32 P-labelled TAR RNA (nucleotides 1–57; 5,000 c.p.m., 1–5 ng) was incubated with wild-type or mutant Tat protein (100 ng) in binding buffer and analysed on a non-denaturing 6% acrylamide gel²⁹. In the competition analysis, 200 ng of either the specific competitor, TAR RNA, or the nonspecific competitor, tRNA, was added to the reaction mix. **b**, Tat proteins were tested for *in vivo* function by the protein/DNA electroporation method³⁰. **c**, Affinity chromatography with wild-type and mutant Tat proteins. The affinity column protocol and *in vitro* transcription assays are described in Fig. 1 legend. **d**, Immunoprecipitation of holo-TFIID with anti-Tat antibody. HeLa holo-TFIID was preincubated with either wild-type or Lys-41 mutant peptide 36–50, followed by addition of 35 S-labelled Tat. Tat-TFIID complexes were immunoprecipitated with anti-Tat antibody. Reactions were electrophoresed and western blotted with anti-TAF p250 antibody²⁰ and 125 I-protein A. Lane 2, reaction contains Tat, TFIID and wild-type peptide 36–50; lane 1, reaction contains Tat, TFIID and mutant peptide 36–50. 68K BSA reacts with the p250 peptide-BSA conjugate antiserum.

METHODS. Tat mutant was made by a double PCR method. The entire Tat coding region was sequenced to confirm the presence of the substitution of Lys 41 to Thr. The mutant Tat protein was overexpressed in *E. coli* and purified as described⁶. For **a**, TAR RNA gel shift was performed by incubation of Tat protein and *in vitro*-generated labelled RNA in binding buffer (10 mM Tris, pH 7.5, 1 mM DTT, 1 mM EDTA, 50 mM NaCl, 0.5 U μ l⁻¹ RNase inhibitor (Promega), 0.09 mg ml⁻¹ BSA and 0.05% (v/v) glycerol)²⁹; final reaction volume 20 μ l. Reactions were incubated at 23 °C for 15 min and loaded onto a pre-run 6% non-denaturing acrylamide gel with 0.5 $\times$ TBE (Tris, pH 8.0, 89 mM; boric acid, 89 mM; EDTA, 2 mM) as running buffer. In **d**, 40 μ l HeLa holo-TFIID (P11, 0.85 M; DEAE, 0.15 M)²⁶ was preincubated with 50 μ g wild-type or mutant Tat peptide 36–50 for 2 h at 4 °C, then with 20 μ l 35 S-labelled Tat (TNT, Promega) at 4 °C for 2 h, after which 5 μ g anti-Tat monoclonal antibody (epitope-mapped to Tat amino acids 6–14 (American Biotechnologies) was added to the reaction mixture. Immune complexes were precipitated with 20 μ l packed protein A plus protein G-agarose (Oncogene Science), separated on a 4–20% gradient SDS-PAGE and western blotted with anti-TAF p250 antibody²⁰ and 125 I-protein A. 125 I and 35 S bands were detected on a PhosphorImager (Molecular Dynamics).



tate the assembly of a more processive elongation complex. Alternatively, Tat may function independently at the level of both transcriptional initiation and elongation. □

Received 21 May; accepted 7 December 1993.

- Berkhout, B., Silverman, R. H. & Jeang, K.-T. *Cell* **59**, 273–282 (1989).
- Laspi, M. F., Rice, A. P. & Matthews, M. B. *Genes Dev.* **4**, 2397–2408 (1990).
- Feinberg, M. B., Baltimore, D. & Frankel, A. D. *Proc. natn. Acad. Sci. U.S.A.* **88**, 4045–4049 (1991).
- Marciniak, R. A. & Sharp, P. A. *EMBO J.* **10**, 4189–4196 (1991).
- Southgate, C. D. & Green, M. R. *Genes Dev.* **5**, 2496–2507 (1991).
- Bohan, C. A. et al. *Gene Exp.* **2**, 391–407 (1992).
- Cullen, B. R. *Microbiol. Rev.* **56**, 375–394 (1992).
- Kato, H. et al. *Genes Dev.* **6**, 655–666 (1992).
- Nakajima, N., Horikoshi, M. & Roeder, R. G. *Molec. cell. Biol.* **8**, 4028–4040 (1988).
- Timmers, H. Th. M. & Sharp, P. A. *Genes Dev.* **5**, 1946–1956 (1991).
- Comai, L., Tanese, N. & Tjian, R. *Cell* **68**, 965–976 (1992).
- Kassavetis, G. A. et al. *Cell* **71**, 1055–1064 (1992).
- Lobo, S. M., Tanaka, M., Sullivan, M. L. & Hernandez, N. *Cell* **71**, 1029–1040 (1992).
- Taggart, A. K. P., Fisher, T. S. & Pugh, B. F. *Cell* **71**, 1015–1028 (1992).
- White, R. J. & Jackson, S. P. *Cell* **71**, 1041–1053 (1992).

- Kuppuswamy, M., Subramanian, T., Srinivasan, A. & Chinnadurai, G. *Nucleic Acids Res.* **17**, 3551–3562 (1989).
- Rice, A. P. & Carlotti, F. J. *Virology* **64**, 1864–1868 (1990).
- Tiley, L. S., Brown, P. H. & Cullen, B. R. *Virology* **178**, 560–567 (1990).
- Boyer, T. G. & Berk, A. J. *Genes Dev.* **7**, 1810–1823 (1992).
- Hisatake, K. et al. *Nature* **362**, 179–181 (1993).
- Ruppert, S., Wang, E. H. & Tjian, R. *Nature* **362**, 175–179 (1993).
- Chiang, C.-M., Ge, H., Wang, Z., Hoffmann, A. & Roeder, R. G. *EMBO J.* **12**, 2749–2762 (1993).
- Kao, C. C. et al. *Science* **248**, 1646–1650 (1990).
- Ha, I., Lane, W. S. & Reinberg, D. *Nature* **352**, 689–695 (1991).
- Formosa, T., Barry, J., Alberts, B. M. & Greenblatt, J. *Meth. Enzym.* **208**, 24–45 (1991).
- Meisterernst, M., Roy, A. L., Lieu, H. M. & Roeder, R. G. *Cell* **66**, 981–993 (1991).
- Maldonado, D., Ha, I., Cortes, P., Weis, L. & Reinberg, D. *Molec. cell. Biol.* **10**, 6335–6347 (1990).
- Helin, K. et al. *Cell* **70**, 337–350 (1992).
- Roy, S., Delling, U., Chen, C.-H., Rosen, C. A. & Sonenberg, N. *Genes Dev.* **4**, 1365–1373 (1990).
- Kashanchi, F., Duval, J. F. & Brady, J. N. *Nucleic Acids Res.* **20**, 4673–4674 (1992).

ACKNOWLEDGEMENTS. We thank A. Tatro and A. Kumar for assistance in the RNA binding studies; R. Tjian and N. Thompson for monoclonal antibodies against TBP; S. Jackson for polyclonal antibodies against SP1; A. Hoffmann for TAF antibodies; A. Berk for the epitope-tagged TBP cell line LTRα; and colleagues in the Laboratory of Molecular Virology for discussion. This work was supported in part by NIH grants to R.G.R. and by awards from the Helen Hay Whitney Foundation and the Aaron Diamond Foundation to C.M.C.

PLEASE follow these guidelines so that your manuscript may be handled expeditiously.

Nature is an international journal covering all the sciences. Contributors should therefore bear in mind those readers who work in other fields and those for whom English is a second language, and write clearly and simply, avoiding unnecessary technical terminology. Space in the journal is limited, making competition for publication severe. Brevity is highly valued. One printed page of *Nature*, without display items, contains about 1,300 words.

Manuscripts are selected for publication according to editorial assessment of their suitability and reports from independent referees. They can be sent to London or Washington and should be addressed to the Editor. Manuscripts may be dealt with in either office, depending on the subject matter, and will where necessary be sent between offices by overnight courier. High priority cannot be given to pre-submission enquiries; in urgent cases they can be made in the form of a one-page fax. All manuscripts are acknowledged on receipt but fewer than half are sent for review. Those that are not reviewed are returned as rapidly as possible so that they may be submitted elsewhere without delay. Contributors may suggest reviewers; limited requests for the exclusion of specific reviewers are usually heeded. Manuscripts are usually sent to two or three reviewers, chosen for their expertise rather than their geographical location. Manuscripts accepted for publication are processed from the London office.

Nature requests authors to deposit sequence and crystallographic data in the databases that exist for this purpose.

Once a manuscript is accepted for publication, contributors will receive proofs in about 4 weeks. *Nature's* staff will edit manuscripts with a view to brevity and clarity, so contributors should check proofs carefully. Manuscripts are generally published 2–3 weeks after receipt of corrected proofs. *Nature* does not exact page charges. Contributors receive a reprint order form with their proofs; reprint orders are processed after the manuscript is published and payment received.

Categories of paper

Review Articles survey recent developments in a field. Most are commissioned but suggestions are welcome in the form of a one-page synopsis addressed to the Reviews Coordinator. Length is negotiable in advance.

Progress articles review particularly topical developments for a nonspecialist readership. They do not exceed 4 pages in length. Suggestions may be made to the Reviews Coordinator in the form of a brief synopsis.

Articles are research reports whose conclusions are of general interest and which are sufficiently rounded to be a substantial advance in understanding. They should not have more than 3,000 words of text (not including figure legends) or more than six display items and should not occupy more than five pages of *Nature*.

Articles start with a heading of 50–80 words written to advertise their content in general terms, to which editors will pay particular attention. The heading does not usually contain numbers, abbreviations or measurements. The introduction to the study is contained in the first two or three paragraphs of the article, which also briefly summarize its results and implications. Articles have fewer than 50 references and may contain a few short subheadings.

Letters to Nature are short reports of outstanding novel findings whose implications are general and important enough to be of interest to those outside the field. Letters should have 1,000 or fewer words of text and four or fewer display items. The first paragraph describes, in not more than 150 words and without the use of abbreviations, the background, rationale and chief conclusions of the study for the particular benefit of non-specialist readers. Letters do not have subheadings and contain fewer than 30 references.

Commentary articles deal with issues in, or arising from, research that are also of interest to readers outside research.

News and Views articles inform nonspecialist readers about new scientific advances, sometimes in the form of a conference report. Most are commissioned but proposals can be made in advance to the News and Views Editor.

Scientific Correspondence is for discussion of topical scientific matters, including those published in *Nature*, and for miscellaneous contributions. Priority is given to letters of fewer than 500 words.

Preparation of manuscripts

All manuscripts should be typed, double-spaced, on one side of the paper only. An original and four copies are required, each accompanied by artwork. If photographs are included, five sets of originals are required; for line drawings, one set of originals and four good-quality photocopies are acceptable. Reference lists, figure legends and tables should all be on separate sheets, all of which should be double-spaced and numbered. Three copies of relevant manuscripts in press or submitted for publication elsewhere should be included with submitted manuscripts, clearly marked as such. Five copies of revised and resubmitted manuscripts, labelled with their manuscript numbers, are required, together with five copies of a letter detailing the changes made.

Titles are brief and simple. Active verbs, numerical values, abbreviations and punctuation are to be avoided. Titles should contain one or two key words for indexing purposes.

Artwork should be marked individually and clearly with the author's name and, when known, the manuscript number. Ideally, no figure should be larger than 28 by 22 cm. Figures with several parts are to be avoided and are permitted only if the parts are closely related, either experimentally or logically. Unlettered originals of photographs should be provided. Suggestions for cover illustrations, with captions and labelled with the manuscript number, are welcome. Original artwork is returned when a manuscript cannot be published.

Protein/nucleotide sequences should ideally be in the three-letter and not the single-letter code for amino acids. One column width of *Nature* can accommodate 20 amino acids or 60 base pairs.

Colour artwork. A charge of £500 per page is made as a contribution towards the cost of reproducing colour figures. Inability to pay these costs will not prevent publication of essential colour figures if the circumstances are explained. Proofs of colour artwork may be sent to contributors under separate cover from their galley proofs.

Figure legends should not exceed 300 words and ideally should be shorter. The figure is described first, then, briefly, the method. Reference to a method published elsewhere is preferable to a full description. Methods are not described in the text.

References are numbered sequentially as they appear in the text, followed by those in tables and finally by those in figure legends. Only papers published or in the press are numbered and included in the reference list. All other forms of reference should be cited in the text as a personal communication, manuscript submitted or in preparation. Text is not included in reference lists. References are abbreviated according to the *World List of Scientific Periodicals* (Butterworths, London, 1963–65). The first and last page numbers are included; reference to books includes publisher, place and date.

Abbreviations, symbols, units and Greek letters should be identified the first time they are used. Acronyms should be avoided whenever possible and, if used, defined. Footnotes are not used in the text.

Acknowledgements are brief and appear after the reference list; grant and contribution numbers are not allowed.

Supplementary information is material relevant to Articles or Letters which cannot, for lack of space, be published in full, but which is available from *Nature* on request.

Submission. Manuscripts can be sent to the Editor at 4 Little Essex Street, London WC2R 3LF, UK or at 1234 National Press Building, Washington, DC 20045, USA. A telephone and fax number should be included. Manuscripts or proofs sent by air courier to London should be declared as 'manuscripts' and 'value \$5' to prevent the imposition of import duty and value-added tax.